THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVI | January, 1925 | Number 1

THE REGRESSION OF SIBLINGS OF CHILDREN WHO TEST AT OR ABOVE 135 IQ (STANFORD-BINET)

FOR THE COMMITTEE*

MARGARET V. COBB AND LETA S. HOLLINGWORTH

Teachers College, Columbia University

The data to be reported under the above title have been derived from the same classes, under the same auspices, as those reported from Public School 165, Manhattan, previously.[5] The 58 pupils testing at or above 135 IQ (Stanford-Binet), for whom we have secured data as to family,† have 54 living siblings.†† The distribution of these, as to size of family, is as follows:

NUMBER OF CHILDREN IN FAMILY	FREQUENCY	NUMBER OF SIBS OF OUR PUPILS
1	18	0
2	28	28
3	9	17[2]
4	3	9
	58	54

* This report is rendered as part of the work of the joint committee in charge of special classes for bright children, previously described elsewhere.[5] The members of this committee are Mr. Jacob Theobald and Miss Jane Monahan, of Public School 165, Manhattan; Miss Margaret V. Cobb, Dr. Grace A. Taylor, and Dr. Leta S. Hollingworth, of Teachers College, Columbia University. The work is being carried out with the advice of District Superintendent John E. Wade, and in cooperation with the Division of Educational Psychology, of the Institute of Educational Research, of Teachers College. The statistical and clerical work of this study was provided for by funds granted through the Institute, by the Carnegie Corporation of New York.

† One child who joined the class for a brief period was withdrawn before we could collect data as to siblings.

†† Two brothers in our group have the same third sibling (an infant), counted once only. Each of these brothers is counted both as pupil and as sibling in the distributions; here each is counted as sib to the other.

Of these 54 siblings, 47 have been tested for intelligence—45 by the present writers, 1 at The Little Red Schoolhouse, and 1 at the College of the city of New York. Four siblings are infants, too young to test. The three adult siblings of one child in our group (IQ 143) did not cooperate with us. Two of these are but half-sibs of our pupil. In compiling our data for intelligence, this child has been omitted from the calculations, leaving 57 pupils to be included. Our data are, therefore, almost perfect, as respects freedom from the fallacies of selection. We have tested every full sibling over three years of age, of every child included in our category of investigation, except one. This sole omission could not affect the trend of our findings in any important way. The category of investigation may be defined as *children in our classes, who were available at the time when inquiry as to sibs was instituted.*

TABLE I.—DISTRIBUTION OF PUPILS TESTING ABOVE 135 IQ, IN COMPARISON WITH DISTRIBUTION OF THE SIBLINGS OF THE GROUP

IQ	All pupils of special class	Siblings	Pupils of special class with tested siblings
190–185	2	0	1
185–180	0	0	0
180–175	1	0	1
175–170	4	2	3
170–165	4	1	4
165–160	5	0	3
160–155	8	0	5
155–150	7	2	4
150–145	7	2	6
145–140	9	1	6
140–135	6	2	3
135–130	4	5	1
130–125	0	4	0
125–120	0	4	0
120–115	0	1	0
115–110	0	3	0
110–105	0	4	0
105–100	0	1	0
100–95	0	2	0
	—	—	—
n =	57	34	37
Average IQ =	153.5	129.2	154.9

As stated, 18 of the 57 children included in our calculations have no living siblings, and two more have infant siblings only; but in order to show a *group* such as ours, against the siblings attached to it, we have presented all of the 57 pupils, distributed against the 34 siblings, who were of an age to have their intelligence measured in terms of IQ (Stanford-Binet). Table I gives this comparison.

The 18 "only" children are distributed nearly at random as to IQ through our group, the mean IQ for pupils with siblings being 154.9 as against the mean of 129.2 for the siblings, and of 150.5 for those who are "only" children. The mean IQ of the 20 children having no siblings, or infant siblings only, and hence not included in the comparison of the third column in Table I, is 151.1. Children having adult* siblings only are included in the third column of Table I, because inspection of Table III indicates that the results for adult siblings would distribute approximately as do those for the younger siblings if they could be translated into terms of IQ.

Table II gives the comparison between siblings of our pupils who test *above* 150 IQ, and the siblings of those who test *below* 150, but at or above 135 IQ.

TABLE II.—SHOWING HOW SIBLINGS OF CHILDREN ABOVE 150 IQ COMPARE IN IQ WITH SIBLINGS OF CHILDREN BETWEEN 135 AND 150 IQ

IQ	SIBS OF CHILDREN 150–188 IQ	SIBS OF CHILDREN 135–150 IQ
175–170	2	0
170–165	0	1
165–160	0	0
160–155	0	0
155–150	2	0
150–145	2	0
145–140	1	0
140–135	0	2
135–130	2	3
130–125	3	1
125–120	2	2
120–115	1	0
115–110	2	1
110–105	3	1
105–100	0	1
100– 95	0	2
	—	—
	n = 20	14
	Average IQ = 132.8	124.0

* By "adult" is here meant, past the fourteenth birthday.

There is a difference of about nine points between the means of the two groups of siblings, in favor of those related to children above 150 IQ. In itself this difference between means is not very reliable, as will be plain from inspection of the variability within each group. However, in the identically constituted comparison of the 13 adult siblings, who could not be reliably measured by Stanford-Binet (as they are very intelligent), but were measured instead by Army Alpha, Form 5, we find a difference in the same direction. Table III gives the results for Army Alpha, showing a difference between means of 22 points, in favor of siblings attached to children testing above 150 IQ.

TABLE III.—SHOWING STATUS OF SIBLINGS PAST THE FOURTEENTH BIRTHDAY, ON ARMY ALPHA, FORM 5, IN TERMS OF POINTS EARNED

Sibs of children testing 150–188 IQ			Sibs of children testing 135–150 IQ		
Sib	Alpha score	Birthday age	Sib	Alpha score	Birthday age
A	185	18–3	H	166	18–4
B	171	20–9	I	153	17–0
C	167	15–10	J	139	14–11
D	159	14–6	K	135	14–4
E	142	18–2	L	130	16–11
F	138	16–0	M	87	20–2
G	134	15–11			
Average....	156.7	17–1	Average...	134.8	16–11

The mean for the first group nearly equals the mean for Yale freshmen, the highest group of freshmen yet reported, as regards Alpha scores; and is about the same as that reported in general for postgraduate students in first-rate universities. The mean for the second group is about that generally reported for college freshmen. The mean of both groups of siblings taken together is about that of college seniors, in first-rate American colleges. The mean ages of the two groups here compared are practically identical, so that none of the difference found may be ascribed to difference in maturity.

On the whole it seems clear that a pupil testing above 150 IQ has a distinctly better chance to be accompanied by a sibling of very high ability, than does a pupil testing from 135 to 150 IQ. All but 4 of the 15 siblings who test at or above 140 IQ or above 150 points on Army Alpha, belong to pupils who themselves test above 150 IQ. On

the other hand, of the 12 siblings testing at or below 117 IQ or below 100 points on Army Alpha, 6 belong to pupils who themselves test above 150 IQ. No correlations have been calculated, as we are dealing here with a severely restricted range of competency.

Finally, it may be stated that the siblings of children who test in the best one-half of 1 per cent for intelligence, are nearly all contained within the upper half of the total range of IQ, so far as our data tell. These siblings show, however, a marked regression toward the mean of the general population, as compared with the fraternal group to which they belong. Galton[3] believed he had established in the case of stature, that "the unknown brother of a known man is probably only two-thirds as exceptional as he is." He placed the value of "fraternal regression," therefore, at two-thirds, which is to say that he found the brothers of exceptional men to be only two-thirds as exceptional as they, on the average, in stature.

Discussing these phenomena, Galton said: "However paradoxical the fact may seem at first, of there being such a thing as Fraternal Regression, a little reflection will show its reasonableness, which will become clearer later on . . . . We may recollect that the unknown brother has two different tendencies, the one to resemble the known man, and the other to resemble his race. The one tendency is to deviate from P (the mean of the population) as much as his brother, and the other tendency is to deviate not at all. The result is a compromise." Also, "It must be a universal rule that the unknown Kinsman in any degree of a known Man, is on the average more mediocre than he is."

The hypothesis last quoted calls for greater precision of meaning, or at least of statement, to secure itself. "Known" men in the sense of "measured" men may, and probably will, unless carefully selected, fall at a mean coinciding with the mean of the general population. There will, in that case, be no fraternal regression, since their brothers' mean will tend to fall at the same point. If the word "deviating" or "exceptional" be substituted for "known" in Galton's statement, the hypothesis becomes thoroughly intelligible; and this must be what Galton intended.

In the case of intellect, with our widely deviating group, the fraternal regression (counting sibs regardless of sex) toward 100 IQ, appears as $\frac{29.2}{54.8}$, or just about three-fifths (in terms of IQ). We may well assume that this close approximation to Galton's finding for stature is probably a mere coincidence. Ratios cannot be determined

with mathematical validity in terms of IQ, since points of IQ are not, or are not known to be, true units. Still the ratio of near three-fifths holds also, if we consider the amount of regression in terms of PE of the distribution. The mean of our highly selected pupils falls at about 7 PE from 100 IQ; that of their siblings falls at near 4 PE from 100 IQ, in the same direction (as 1 PE = ± 8 IQ, according to Terman's distribution[2]). Moreover, when individual tests have been used as distinguished from group tests, correlations to determine resemblance between siblings, as regards intelligence yield, on the whole, coefficients clustering between .50 and .60, for full siblings; and of about .80 for those who are twins.[10] (None of our pupils is a twin.) Rough group tests yield much lower coefficients.

Taking these various facts into consideration, it is suggested that the unknown siblings of a group of extremely intelligent children will be found about three-fifths as exceptional, in the same direction, as the selected children in terms of the distribution of IQ. We now require, in this connection, determinations of fraternal regression when those selected to form the starting point of investigation have a mean at 130, at 120 and at 110, for instance.

It is of some interest, also, that Galton at first determined the value of filial regression (the regression of sons from mid-parent) at three-fifths for stature, later fixing upon two-thirds for both filial and fraternal regression.

The scholastic achievement of the siblings here studied deserves comment. For 19 of them who have been in school for six years or longer, the median number of grades "skipped" is three. All but one have "skipped" at least once. These siblings have saved to the public schools many years of school time, and many hundreds of dollars, by their unusual competency. The only cases of repeating are of two siblings who repeated one grade (and later "skipped" one grade each).

Of the 47 siblings tested, 25 are sisters and 22 are brothers. The groups are too small to make comparisons of brothers with sisters profitable as the total group, small in itself, must be broken into four sub-groups for such comparison. The following facts are stated as of possible incidental interest.

The mean IQ of the 19 sisters tested by Stanford-Binet is 124.8. Of the 15 brothers so tested, the mean IQ is 134.7. The mean score in points of the 6 adult sisters, tested by Army Alpha is 150.0. Of the 7 adult brothers so tested the mean score in points is 139.3. Our pupils who test above 150 IQ have 16 sisters and 12 brothers. The two highest scores made on Army Alpha were scores of sisters.

The highest IQ is also that of a sister (173), and the next to highest, that of a brother (172).

The siblings of the very bright children of our group are few, but they are nearly all living. The data as regards percentage of death indicates that about 95 per cent of the children born into these families are alive, there being now 112 in all. Of the 57[3] families studied 35 are complete by the criterion often adopted, nine years having elapsed since the birth of a child. In at least three families there can be no more full brothers or sisters, since one parent is dead. There will probably be additional siblings born in the other families, but not many. The contrast in this respect with the preceding generation is very pronounced. Our pupils average a little less than one sibling each. Their parents had six or seven times as many brothers and sisters, as present data from family histories indicate uncles and aunts in about that proportion.

Thus the study of the siblings of very gifted children, born within the latest decade and now about ten years old, proceeds somewhat slowly because so many of these children have no siblings. Our pupils have about 200 first cousins, who may be tested. Thirty-three of these tests made to date yield a mean IQ of 127. These doubtless represent a favorable selection of cousins, so that the result may be expected to change, as we increase our sample. Mental tests have at this date been made, also, in the case of 20 parents, using Army Alpha. They yield a median of 142 points, which is about the median score for college seniors in this country.

References

1. Elderton, E. M.: A Summary of the Present Position with Regard to the Inheritance of Intelligence. *Biometrika*, 1923.
2. Galton, F.: "Hereditary Genius." Macmillan Co., London, 1869.
3. Galton, F.: "Natural Inheritance." Macmillan Co., London, 1889.
4. Gordon, K.: "The Influence of Heredity on Mental Ability. Report to the State Board of Control," Sacramento, Cal., 1921.
5. "National Society for the Study of Education—Twenty-third Year Book." Public School Publishing Co., Bloomington, Ill., 1924.
6. Pearson, K.: Comparison of Association in Intelligence of Pairs of Siblings as Determined by Broad Categories and Binet-Simon Tests. *Biometrika*, 1918.
7. Pintner, R.: The Mental Indices of Siblings. *Psychological Review*, 1918.
8. Schuster, E. and Elderton, E. M.: The Inheritance of Ability. *Eugenics Laboratory Memoirs I*, London, 1907.
9. Terman, L. M.: "The Measurement of Intelligence." Houghton Mifflin Co., Boston, 1916.
10. Thorndike, E. L.: Measurements of Twins. *Columbia University Contributions to Philosophy and Psychology*, 1905.

HIGH TEST SCORES ATTAINED BY SUB-AVERAGE MINDS

OSCAR W. RICHARDS

University of Oregon

AND

S. C. KOHS

Director Jewish Federation, Oakland, Calif.

The continued and increasing utilization of the alternative type test for purposes of determining mental and pedagogical progress has inclined the writers to make emphatic the caution that this method is subject to a high probable error; and that conclusions based upon a test of this character, alone, are likely to be invalid and misleading to an unusual degree.

This article presents some theoretical considerations with regard to the use of this type of test and empirical evidence which indicates its inadequacy when used alone. For convenience the article will be divided into three parts. The first part will partly answer and criticize certain questions brought up in two recent articles. The second part will test experimentally the conclusions previously reached by Kohs. Part three contains a general analysis of chance as it affects test-construction and test-application.

PART I

West[9] recently gave a group of true-false tests and plotted the frequencies of his scores against the normal law of error. Because his data did not conform to the so-called normal law he concludes that there was something wrong with the right-minus-wrong method of scoring. His article is an excellent example of what Boring[2] has termed the normal law fetish.

When referring any set of experimental data to any law we should first determine whether or not the law is applicable to the particular case. The graph of the normal or Gauss-Laplace Law of error

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

is a symmetrical curve with a maximum at the origin and extending to x equals plus infinity and to x equals minus infinity. All points on the curve have actual values. West's data start, provided he had

not used an inadequate method of scoring (to be discussed later), from zero at the origin with a maximum value at approximately one-half the number of cases ($N \times p$). His curve then again reaches the x-axis at a definite number N. His values have meaning only for the numbers in his table, that is there is no ordinate value to correspond with $x = 7.71$. Clearly the normal law is inadequate. The curve that West should have used is the binomial law

$$P = \frac{n!}{r!(n-r)!} p^r q^{n-r}$$

where P is the probability of the pupils guessing r items right and $n - r$ items wrong with a true-false test having n items. For the alternative test $p = \frac{1}{2} = q$. This curve starts at the origin rising to a maximum at np and falling to the x-axis again at n. This curve is symmetrical only when the probability for each item is one-half.

It can be shown geometrically that the normal curve is the limitating case of the binomial when n is infinitely large and the binomial formula can be converted into a formula very similar to that of the normal law by the use of Stirling's Formula for large factorials. However, the derived formula would be of no use to us unless we had at least 1000 items because Stirling's Formula does not give accurate values for smaller numbers. When educational measuring is affected by so many highly variable factors as attention, common sense clues, etc., it would seem that the statistical analysis of it should be done only by the most accurate and adequate methods rather than by taking some much used formula with the hope that by some remote chance it is the correct and adequate formula. Certainly no field needs an understanding of mathematical statistics more than the educational one. No field has more inadequate textbooks.

In 1920, Kohs[6] pointed out from a theoretical viewpoint that the right-minus-wrong scoring method was inadequate. In another place Richards[8] has indicated ways and means for making more useful scoring scales for the multiple response test that are free from certain objections. This last paper gives a discussion of the binomial law in a form well adapted for educational use and extends the theoretical treatment to include the three-and-four alternative tests. The measuring value of these tests as affected by chance is discussed. The writer wishes to call the attention of those educators who will use these formulas to Glover's[5] recent tables. These tables are an almost invaluable aid. While the illustrative material is from the

medical and biological field, Pearl's[7a] new book has much information of value to the nonmathematical educational reader. The educator with a moderate mathematical knowledge might read Fisher[3] with profit.

Bateson's[1] article raises certain questions with regard to the arrangement and sequences of right and wrong items in this type test. His statements made apparently without any data to prove them are in the main right. Certain of these will be discussed in a more adequate manner in Part III of this paper. Bateson indicates that McCall[7] has proved without any doubt that the right-minus-wrong method entirely does away with the factor of guessing. As was previously said, Kohs[6] indicated that this method did not rest on sound logic.

We may well reëxamine McCall. He tossed a penny several thousand times and obtained *heads* only one more time than he obtained *tails*. This meant that he had a remarkably unbiased penny. From this he concludes that a child guessing how to mark the items on a test will obtain as many wrong as right. Hence if we subtract the number wrong from the number right the score will be zero as it should be. However, had McCall tossed 50 pennies at a time he would have found that the special combination of 25 *heads and* 25 *tails* would have occurred only about *once* in every *nine* throws. Consequently this method of scoring is accurate only once in every nine cases. The other cases would get in the main small plus and minus scores. For those who missed Kohs' article it will be summarized in Part II. It is hoped that the present article may in part dispel this erroneous conclusion. The educator who still believes in this method is advised to toss pennies in groups of from 20 to 50 or refer to any good discussion of the binomial law.

Part II

In the previous paper by Kohs it was indicated that if a child was given a true-false test consisting of 50 items, and proceeded to mark the items plus or minus, knowing nothing of their content, but guessing at each in order, that instead of obtaining 25 right and 25 wrong, or a zero score, most frequently he would in reality obtain such a score only *once* in *nine* chances. It was further indicated, that were many children to be tested of those guessing, approximately 45 per cent would obtain positive scores and the other 55 per cent would obtain negative and zero scores. The mathematical calculations also seemed to point

to the fact that about 11 per cent would obtain positive scores of a value over 16 per cent and practically none would obtain scores higher than 28 per cent. The accuracy of the figures would, of course, be more apparent the larger the number examined. The following experiment was designed to test these conclusions.

Mimeographed lists of 26, 40, and 50 numbers arranged in vertical columns were printed, and 52 of these sheets were given to the students of the statistics class at Reed College in the spring of 1919. The students were instructed to mark each column so that there would be an equal number of plus and minus signs indiscriminately placed. These sheets were later scored with a key constructed by selecting a random group of numbers from the Oregon Key List for the selective draft, calling the even numbers plus and the odd numbers minus. This yielded a key of purely chance order. The score for each sheet was stated in terms of percentage, found by subtracting the number wrong from the number right and multiplying the difference by two. The columns of 50 were naturally used without change. The 26 columns were paired by combining two papers, the twenty-sixth item of one having been marked plus and the twenty-sixth item of the other having been marked minus. After discarding the twenty-sixth item of each, an additional series of 50 was thus made available.

In order to bring the total number of sheets to 100 the 40 columns of each of 22 of these papers was extended to 50 by adding 10 items. Eleven of these 22 papers were then marked in a purely chance order by Kohs and the other eleven by Richards. These were then scored in the same fashion as the preceding material.

TABLE I.—FREQUENCY OF SCORES IN EXPERIMENTAL GROUP

Per cent	Theoretical expectancy	Actual + scores	Actual − scores	Deviation	
				+ Scores	− Scores
0	11	5	..	−6	
4	11	14	14	+3	+3
8	10	6	2	−4	−8
12	8	10	8	+2	+0
16	6	11	7	+5	+1
20	4	5	6	+1	+2
24	3	1	1	−2	−2
28	2	2	1	+0	−1
32	1	1	2	+0	+1
40	0.2	3	1	+2.8	+0.8

	Theoretical Expectancy	Actual Result
Positive scores	45	53
Negative and zero scores	55	47
Positive scores above 16 per cent	11	11
Positive scores above 28 per cent	0	4

Table I presents the tabulation of the results first comparing them with theoretical frequency and second indicating the deviation from each expected frequency.

The results, as a whole, do not deviate greatly from the normal expectation, and a closer approximation to the theoretical probability would have been obtained had the empirical material been greatly increased. However, certain discrepancies are significant. Theoretically one would expect a zero score to occur once in *nine* times but actually it occurred only *once in twenty*. Instead of *eleven* zero scores, based on the theoretical prediction, there were actually only *five*. Theoretically 45 per cent of the cases should be positive and 55 per cent negative and zero scores. Instead, 53 per cent positive scores and 47 per cent zero and negative scores were actually obtained experimentally. Of the positive scores, 11 per cent should, under normal circumstances, be above 16 per cent and this is borne out by the obtained data, the result tallying perfectly with expectation in this instance. Almost no positive scores were expected above 28 per cent, yet it is very significant that *three* of the scores were 40 per cent.

The results of both the experimental and the theoretical studies indicate definitely that "*chance takes care of the guesser*" *most frequently to the guesser's advantage*, and that the possibility of obtaining significant positive scores are about two out of five while the chance of securing a zero score is not likely to occur more often than about once in nine times.

One of the reasons for the slight inconsistency between the distributions of theoretical expectancy and those experimentally determined was the personal bias of the students who filled the blanks. Some students contributed as many as six papers and in extending the 40 columns both writers supplemented the lists on 11 papers each. In order to eliminate this human personal equation in setting down plus and minus signs Part III of this paper presents a study of probability where only pure chance operated. These results have been compared with theoretical figures of the original study and also with the results of the experiment. The values in the column for theoretical expecta-

tion have been rounded for convenience. This, of course, will not materially effect the comparisons and conclusions. The more accurate figures may be found in Kohs[6] article.

Part III

The numbers found in the lists of the selective draft for Oregon were used to study empirically chance order and how chance occurrences would agree with normal expectation. This list was chosen because it is probably the best available example of numbers arranged purely by chance, guaranteed both as to accuracy and to secrecy because of strict governmental control. When men's lives are subject to call for the supreme sacrifice one may be certain all possible precautions have been taken to secure an unbiased working of pure chance. This list, therefore, of 6649 numbers was taken by fifties consecutively and charted on graph paper. The even numbers were recorded plus and the odd numbers minus. In order to secure the full quota of the final series of 50, the last 49 numbers in the list were extended to 50 by tossing up a nickel, calling a possible head plus, a possible tail minus, and recording the result. This yielded a total of 133 series of 50 numbers each. These series were scored, using the first 50 as a key, and the percentages recorded as in Part II, using the formula two times right minus wrong.

No one can state dogmatically what the relative frequencies in the straight runs of two plusses, three, four, or five plusses (or minuses) should be in a chance series of 50 plus and minus terms in a two-alternative type test. The usual arm-chair statistician would deny, possibly, that a run of four plusses or minuses occurred theoretically with sufficient frequency to warrant inclusion in a list of, say, 50 or 100 items. The frequencies of the sequence of plusses and minuses were therefore next recorded. These frequencies are shown for each group of approximately one thousand in Table II.

In the last column is given the probability of each occurrence. A run of 13 plusses occurred only once in the 6000 numbers, hence it is questionable to place 13 consecutive questions to be marked plus, in one test of 50 items. But in devising several tests each of 50 items it would be desirable to have a few runs of four or five plusses or minuses to approach more accurately chance order. By consulting Table III one may find how many of the different length runs should be included in a test of the two-alternative type. Table III merely translates the

probabilities of Table II into actual frequencies. The possibility of error through bias or personal equation is always possible when one merely *guesses* that he is putting the plus and minus items in a purely chance order. It cannot be emphasized too strongly that in devising and utilizing tests of this nature which are inevitably affected by

TABLE II.—FREQUENCIES OF SEQUENCES OF PLUSSES AND MINUSES IN CHANCE ORDER

	1 1000	1001 2008	2009 3008	3009 4001	4002 5001	5002 6000	Total	Prob- ability
1+	133	129	117	124	131	145	779	0.12983
1−	141	129	122	135	120	150	797	0.13283
2+	60	66	65	73	52	52	368	0.06133
2−	51	63	70	71	56	55	366	0.06100
3+	29	25	24	34	31	32	175	0.02916
3−	34	32	34	33	36	25	194	0.03233
4+	12	21	23	12	21	18	107	0.01783
4−	23	17	8	12	16	17	93	0.01550
5+	5	10	8	8	7	9	47	0.00783
5−	7	9	3	5	12	9	45	0.00750
6+	4	7	3	1	3	2	20	0.00333
6−	4	6	5	2	4	2	23	0.00383
7+	5	0	0	3	0	2	10	0.00166
7−	1	2	2	4	2	1	12	0.00200
8+	1	0	1	2	0	1	5	0.00083
8−	1	0	2	0	1	1	5	0.00083
9+	0	1	1	1	0	1	4	0.00066
9−	0	0	2	1	0	0	3	0.00050
10+	0	0	0	1	0	0	1	0.00016
10−	0	0	0	0	0	0	0	0.00000
11+	1	0	0	0	0	0	1	0.00016
11−	0	0	0	0	0	0	0	0.00000
12+	0	0	0	0	0	0	0	0.00000
12−	0	0	0	0	0	0	0	0.00000
13+	1	0	0	0	0	0	1	0.00016
13−	0	0	0	0	0	0	0	0.00000

chance as well as by other factors such as adaptation or common sense clues, no pains should be spared to make the mechanism as accurate as is humanly possible.

In Table IV an analysis of the 132 series of 50 numbers is found. It will be observed that the results here conform more closely to theoret-

ical expectation both as to frequency of zero scores and high scores than was the case of the empirical data. However, we find 14 instead of 15 zero scores. This indicates that to obtain the expected number of zero scores at least 200 tests must be given, and that the fewer the number tested naturally the fewer the number of zero scores obtainable. Consequently in the case of a few series it is found that less than the expected zero scores occur. Two less positive scores and two more zero scores are obtained which balance the opposite discrepancy of Part II, and tend to prove that the more cases are examined the closer can one approach the theoretical expectancy presented by Kohs[6] in his original paper. There are three positive scores above 16 per cent more than expected, and incidently one score of 32 per cent occurred as was expected. It may be of interest to note that in this series of 132 cases no positive score above 32 per cent occurred, while in Part II three such scores occurred. However, this does not prove necessarily that such high scores may not occur again in another sampling, but it does indicate that their occurrence is infrequent in purely chance arrays of figures.

TABLE III.—FREQUENCIES OF SEQUENCES OF PLUSSES AND MINUSES TO BE INCLUDED IN A GIVEN TEST

Items	26	50	74	100	150	200
1+	4	7	12	14	20	26
1−	4	7	12	14	20	26
2+	2	4	6	7	10	12
2−	2	4	6	7	10	12
3+	1	2	3	3	5	6
3−	1	2	3	3	5	6
4+	*	1	1	2	3	4
4−	*	1	1	2	3	4
5+	..	..	..	1	1	2
5−	..	..	..	1	1	2
6+	..	..	..	..	*	1
6−	..	..	..	..	*	1

* Indicates only one run of either + or −.

In conclusion, this analysis of data in which the order of numbers is unquestionably of a chance variety proves definitely that the possibility of small positive scores occurring must not be ignored in evaluat-

ing and understanding the results of the two-alternative type of test. Such deviations from expected zero scores may occur more frequently in the testing of border-zone cases and in the upper group of feeble-minded where the subjects are more inclined to guess what they do

TABLE IV.—FREQUENCY OF SCORES IN CHANCE ORDER GROUP

Per cent	Theoretical expectancy	Actual + scores	Actual − scores	Deviation	
				+ Scores	− Scores
0	15	14	..	−1	
4	15	13	15	−2	±0
8	13	5	13	−8	±0
12	11	8	15	−3	+4
16	8	13	5	+5	−3
20	5	10	6	+5	+1
24	4	4	3	±0	−1
28	3	3	1	±0	−2
32	1	1	1	±0	±0
36	1	0	2	−1	+1

	Theoretical Expectancy	Actual Result
Positive scores	59	57
Negative and zero scores	73	75
Positive scores above 16 per cent	15	18
Positive scores above 28 per cent	0	1

not know. This is also an important factor when the subjects are told to guess if they do not know. The ever-present chance of the subaverage, either in mentality or educational ability, obtaining a high score due to the mechanics of chance does not invalidate this useful test, but it does conclusively show that this type of test is inadequate when used alone, or when the series of items is a short one.

Many educators are advocating that this type of test be used in educational measurement. In fact, there have been not a few who have advocated that it be used to replace entirely the usual examinations. It is hoped that this study will sufficiently indicate that these tests alone are not an adequate criterion for the passing or the failing of students in school subjects. For example, a child who is given a 50 item true-false test, and happens to know only 25 of the items, then guesses the other 25 may, one chance out of two, mark 15 of these

right and 10 wrong which would give him a score of 60 per cent. This might be no lower than the average of the class. This situation may account for the fact that Gates[4] obtained such low correlation coefficients in his study to determine the practical value of these tests. This operation of the laws of chance in guessing greatly impairs the diagnostic utility of this type of test when used alone, and demonstrates that extreme care must be used in interpreting results derived therefrom, even when used in combination with other tests.

Summary

An analysis of the laws of chance as they apply to the two-alternative type test reveals the possibility of large errors in final scores where subjects guess whether an item in a test-exercise, is true or false. To obviate this difficulty it is recommended that the numbers of items be large, possibly 75 or 100 and that tests of another type be used along with this type. An empirical testing of this type of test was compared with the theoretical expectancy and while the data in general conformed certain differences were significant. The sequences of a large number of numbers were analyzed and agreed with the normal expectation. From this analysis certain suggestions are made with regard to making and scoring of tests.

Bibliography

1. Bateson, W. H.: Reliability of the True-false Form of Examination. *Educational Administration and Supervision*, Vol. X, 1924, pp. 95–103.
2. Boring, E. G.: The Logic of the Normal Law of Error in Mental Measuring. *American Journal Psychology*, Vol. XXXI, 1920, pp. 1–33.
3. Fisher, A.: "Mathematical Theory of Probabilities," Macmillan, 1922.
4. Gates, A. I.: The True-false Test as a Measure of Achievement in College Courses. *Journal Educational Psychology*, Vol. XII, 1921, pp. 276–88.
5. Glover, J. W.: "Tables of Applied Mathematics," Vol. XII, Wahr, 1923.
6. Kohs, S. C.: "High Test Scores Attained by Subaverage Minds. *Psychology Bulletin*, Vol. XVII, 1920, pp. 1–5.
7. McCall, W. A.: "How to Measure in Education," Macmillan, 1922.

7*a*. Pearl, R.: "Medical Biometry and Statistics," Saunders, 1923.

8. Richards, O. W.: High Test Scores Attained by Subaverage Minds. *Journal Experimental Psychology*, Vol. VII, 1924, pp. 148–56.
9. West, P. V.: A Critical Study of the Right-minus-wrong Method. *Journal Educational Research*, Vol. VIII, 1923, pp. 1–10.

MEMORY BY VISUAL AND BY AUDITORY PRESENTATION[1]

D. A. WORCESTER

Ohio State University

In recent years an unusual amount of interest has been attached to the subject of visual education. Training in visual perception now has a very large place in our elementary schools. This emphasis is without doubt correctly placed, yet there are some who remind us that the average non-professional person, as well as the professional man himself, still secures a large amount of his information from conversations, addresses, sermons, etc. and that nearly all of us are called upon to remember accurately oral directions, orders, business transactions and the like. So many impressions come to us through the ear that it seems hardly probable that that organ has deteriorated as an organ of learning as much as some have feared.

In any event, it is desirable to investigate the efficiency of learning "by ear"—to give the ear a hearing as it were.

Studies as to the relative value of memory by auditory and visual presentation have been made by Pohlmann,[2] Meumann,[3] Münstberg and Bingham,[4] Kirkpatrick,[5] Whitehead,[6] Calkins,[7] Gates,[8] Bennett[9] and others. There is not space here for a review of these researches, but there was no agreement among their findings. In almost all of the experiments the test was of immediate memory alone; the work was done usually with disconnected material, sometimes non-sense materials, and in nearly every case only group results were recorded. In several cases faulty technique invalidated results. For example, to show words on a screen for two seconds each and to speak words at the rate of one word each two seconds does not constitute equal presentation to the two senses involved. During an interval of two seconds one

---

[1] The material of this study was collected in the laboratory of the University of Colorado.

[2] "*Beitrage zur Lehre von Gedächtness.*" 1906, pp. 142–144.

[3] "The Psychology of Learning."

[4] *Psychology Review*, 1894, Vol. I, pp. 34–38.

[5] *Psychology Review*, pp. 602–609.

[6] *Psychology Review*, 1896, Vol. III, pp. 258–269.

[7] *Psychology Review*, 1898, Vol. V, pp. 451–456.

[8] *Journal of Experimental Psychology*, 1916, Vol. I, pp. 373–403.

[9] *Journal of Experimental Psychology*, 1916, Vol. I, pp. 404–418.

may re-read a word several times, but there is not opportunity for re-stimulation of the word presented orally.

It is evident that further work is needed on the problem, particularly from the practical work-a-day standpoint. Few of us have common occasion to remember nonsense or disconnected materials. Still fewer are compelled to learn at a stated and uniform speed and almost none are required to reproduce material, as did Ebbinghause at the same rate as it was presented for learning.

It is important, however, to know whether we remember better directions read to us than those we read ourselves; whether the President's message to Congress heard by us sticks better than if read in the newspaper. It is tremendously significant to know if such tests as the Buckingham-Stevenson Test of Place Geography, which is presented orally, can be safely extended in principle; or in these days of "schools of the air" and radio market reports, how much dependence can be put on the report of the program broadcasted by W. D. A. F.

The present study does not fully answer any of these questions, but it is an attempt in that direction.

The problem here discussed is: (1) By which method of presentation, visual or auditory, of meaningful, connected material is learning most easily accomplished, that is, with the fewest repetitions and in the least time; (2) consequent to which of the above methods is the material best retained after lapses of one day, two days, and seven days.

The subjects of the experiment were 13 summer session students of the University of Colorado. Of these 10 were women and three were men, and with the exception of one man all were engaged in public school work, ranging from a kindergarten teacher to a city superintendent. The ages of the subjects varied from 20 to 58 years. They worked from 40 to 70 minutes a day, six days a week. The subjects, with the exception of the writer, were volunteers from a class in Educational Psychology, who worked without credit.

The materials were twenty 100-word prose selections from the works of Arnold and Huxley. While they were not standardized for difficulty, the method of using them and the number used probably counteracted any slight differences in the ease of learning. Because of various circumstances, for example differences in rapidity of learning or of recall, the subjects did not do equal amounts of work, but the selections were taken up in uniform order to whatever point was attained.

The auditory presentation was used first with six of the subjects and the visual presentation was used first with seven of the subjects.

It was desired that conditions of the experiment be as nearly as possible those of ordinary learning and to this end an effort was made to avoid artificiality in the learning methods. It was recognized that this freedom of control might be a source of errors but it was believed that any error because of this would be no greater than one likely to arise from unnatural conditions of learning.

On one occasion the subject was handed a slip of printed material which was to be read over and over until he judged himself able to repeat the subject-matter exactly. On the alternate days the experimenter read a selection aloud until in the judgment of the subject the selection could be repeated exactly. The recitations were oral.

Only one selection was presented at a sitting.

The whole method was always used and a careful record was kept of the number of repetitions required for learning. In case the subject made an unsuccessful attempt at recitation (as often happened at first but rarely later), that part of the selection remaining from the point of failure was read and an additional repetition was recorded. It may be objected that this attempt at recitation was an aid to learning but as a matter of fact the confusion attendant upon the failure seemed to retard rather than to accelerate the process of memorizing. Also, whereas the attempt at recitation was included in the learning time, an unsuccessful attempt at reproduction inevitably made for a poor score. And finally, Katzaroff has shown[1] that a single recitation among several repetitions has little or no effect on the learning time.

The rate of reading was not predetermined but each subject was allowed to receive the presentation at that speed which seemed to him desirable. Unless requested by the subject to do otherwise the experimenter read aloud at a rate of two words per second, that is, a 100 word selection was read in 50 seconds. Frequently, a subject would request that the rate of reading be increased as the memorizing became more nearly complete. This increased speed seemed to be justifiable as it is unquestionably the natural method.

The subjects were allowed to pause for a short time between repetitions in order that they might test "the feeling of security" of the learning, although they were not supposed to systematically review

---

[1] Katzaroff: *Archives de Psychologie*, Vol. VII, 1908, p. 225ff., "La rôle de la recitation comme facteur de la memorization."

the selection. It is possible that this pause should not have been allowed, but whereas it was common to both methods of presentation and whereas the problem was that of comparative scores, it would not seem to affect the results.

With the exception of the writer, each subject learned a set of nonsense syllables by each method of presentation before commencing the experiment proper and, in most cases, other similar sets were memorized in the closing days of the experiment. The objects in giving the first two days to the memorizing of nonsense syllables were (1) to get the subjects into the attitude of the experiment, and (2) to shed a little light perhaps on the old problem of the transfer of training. This second point will not be discussed in this article.

No practice exercises were used. The nature of the experiment did not seem to make them necessary and, in the second place, it is not customary for the average student to use such exercises before undertaking to memorize something.

The subjects were asked not to attempt a reproduction of the material until practically sure that it could be given exactly. As a matter of fact, it was finally decided to accept and record a recitation which did not fall below 95 per cent of perfection. Learning was to be carried only to the point of one successful reproduction although the subjects of course knew that their memory was to be tested after certain lapses of time.

Only one subject worked at a time and that at the same hour every day throughout the experiment.

Tests of the amount retained were made after one day, two days, and seven days from the time of learning.

Times were carefully kept to the even second with a stop watch.

The method of scoring for accuracy was simple and perhaps somewhat arbitrary. The score was the number of correct words reproduced in order. A word added, substituted, or transposed was scored as omitted. No half credits or credits for correct ideas were given. In case of transposition of a word or phrase the element or elements first put out of their order were deducted and the others allowed. For example, if one said "if nevertheless" for "nevertheless if" 1 per cent was deducted. In case a long phrase should be transposed it was possible in this way to lose a large amount of credit although every word of the selection was given. When changes in number or tense were made and the following constructions then made to agree with these changes, only the original errors were recorded.

No selection was started for a subject nor any aid given other than sometimes to answer the question, "Is this the one about ————?," in regard to a selection learned several days previously.

For reproduction subsequent to the day of learning the subject was given as much time as desired in the attempt at recall. This sometimes amounted to 10 or 15 minutes although the greater part of matter was usually recalled within two minutes. At the time of learning the attempt of reproduction was also unlimited unless the subject seemed definitely unable to proceed or had exceeded the 5 per cent of error allowed for a record.

The retained members method of testing their memory was chosen because the amount retained was frequently so large that the relearning method would show retention to be much less than it really was. It would be impossible, *e.g.*, to make a score of 95 per cent to 100 per cent by the relearning method, on the basis of the number of repetitions, when that number was 10 or less. Scores as high as these were often made, however.

All subjects were asked to refrain from reviewing material memorized until the completion of the experiment, that is, until the seven-day record had been taken. In no case did a subject see or hear a selection between the time of learning and the completion of the experiment.

Results

The results of this research can be best shown by means of tables.

For convenience of the reader the subjects are ranked in all of these tables in order of the best records by auditory presentations.

It is to be noted that of the 13 subjects six (A, C, D, E, F, G) learned more rapidly by hearing the matter read to them and seven (B, H, I, J, K, L, M), by reading it themselves.

Little connection is shown between age or sex and rate of learning.

Of the 12 subjects represented in this table five (A, C, D, G, L) memorized the material with fewer repetitions by the auditory method and seven (E, F, H, I, J, K, M) memorized the material with fewer repetitions by the visual method of presentation.

With the exception of three individuals (E, F, L), those subjects which learned in less time by a particular method learned with fewer repetitions by the same method. This was, of course, to be expected and as, of these three, two (E, F) varied in one direction and one (L) in the other, these exceptions probably have no significance.

TABLE I.—COMPARATIVE LEARNING TIMES BY AUDITORY AND VISUAL PRESENTATIONS

Subject	Sex	Age	Average learning time in minutes		Average per cent accuracy at time of learning	
			Auditory	Visual	Auditory	Visual
A	F	21	9.41	11.06	99	99
B	M	26	13.03	6.75	98.5	98.7
C	F	25	16.5	16.86	99.2	99.6
D	F	58	17.07	21.05	99	97
E	F	28	17.08	17.97	98.7	97.1
F	F	42	19.03	22.13	97.7	97.09
G	M	34	21.14	26.41	98.5	100
H	F	34	21.86	11.01	98.75	99
I	F	25	24.4	22.88	97.6	98
J	F	20	24.92	23.85	96.6	98.3
K	M	21	25.93	18.85	98.3	98.6
L	F	43	33.8	33.65	99.75	98.75
M	F	30	37.54	34.14	97.2	97.5

TABLE II.—COMPARATIVE NUMBER OF REPETITIONS REQUIRED TO MEMORIZE BY AUDITORY AND BY VISUAL PRESENTATIONS

Subject	Sex	Age	Average number of repetitions	
			Auditory	Visual
A	F	21	7.75	9.07
B[1]	M	26		
J	F	20	10.7	9
D	F	58	11.33	19
C	F	25	12.8	17.2
F	F	42	14.8	13.8
G	M	34	15.5	25.5
I	F	25	16.2	15.38
K	M	21	16.3	14.25
E	F	28	17.16	14.57
H	F	34	17.25	10
L	F	43	20.75	26
M	F	30	27.6	21

[1] No record was kept of the number of repetitions for subject B.

Again there seems to be no marked relation between age or sex and learning ability.

From the above two tables it would appear that so far as this experiment goes, there exists no general superiority of either of the two methods of learning used, over the other one.

TABLE III.—COMPARATIVE AMOUNTS RETAINED FROM THE AUDITORY AND FROM THE VISUAL METHOD OF PRESENTATION AFTER ONE DAY, TWO DAYS, AND SEVEN DAYS

Subject	Sex	Age	Average per cent accuracy after					
			One day		Two days		Seven days	
			Auditory	Visual	Auditory	Visual	Auditory	Visual
J	F	20	96.33	94.33	96.	94.	96.	95.
A	F	21	95.54	95.07	96.27	96.9	95.65	95.17
B	M	26	92.7	71.78	92.8	79.16	91.79	79.11
C	F	25	91.2	84.6	90.4	85.2	90.	79.4
M	F	30	87.6	97	86.2	96.5	76.8	69.5
F	F	42	86.29	73.63	86.4	74.7	81.85	67.2
H	F	34	81.45	74.5	83	78.5	49	69.
L	F	43	79.75	71.25	79.15	70.75	57.5	41.
I	F	25	79.2	71.	79.8	69.5	72.6	60.63
E	F	28	76.5	78.8	78.5	76.	51	49.4
D	F	58	74.6	61.	72.66	58.	75	59.66
G	M	34	64.5	17.	63.5	21.	63	16.
K	M	21	63.85	63.4	61.83	54.85	45.83	39.3

There are several interesting things shown by this table.

Eleven out of 13 subjects (after seven days 12 out of 13) remembered better the material heard than that seen. Of the two subjects who retained better from the visual method, Subject E required a slightly longer learning time, though less repetitions, by that method. Subject M presents a rather interesting result. Although requiring much more learning time and more repetitions by the auditory method of presentation, retention was markedly better from the visual method. This is to be explained on the ground that the auditory method was so much harder for this subject that she never felt really sure of material acquired in that way. On several occasions, she remarked, "My, that was hard. I don't know whether I will remember it or not."

Summarizing the results of the tables it will be seen that:

Of the six subjects who learned more rapidly through auditory presentation five also retained better from auditory presentation.

Of the five subjects who learned with fewer repetitions through auditory presentation five also retained better from auditory presentation.

Of the seven subjects who learned more rapidly through visual presentation six retained better from auditory presentation.

Of the seven subjects who learned with fewer repetitions through visual presentation five retained better from auditory presentation, with one doubtful, Subject E.

The degree of retention shown in this table is surprisingly high. This was due undoubtedly to the fact that attention was at a high pitch, that the material was completely learned and that but one selection was learned at a sitting.

In general, it would appear that there is an intrinsic superiority for retention in the auditory method of presentation. If this be so, it is of very great importance. It will mean, for instance, that a teacher wishing children to memorize a poem, will do better to read it aloud to them once a day, than to have them read it to themselves once a day. It is probable also that the lecture is more efficacious than the single reading of printed matter and, still further, it implies that it is economical for two students to study orally together.

The age factor seems to have more bearing upon retention than upon learning, although there is no constant relation apparent. It is probable that the age differences would have been more marked in younger subjects. A few interesting results may be noted.

The youngest subject, J, age 20, ranked tenth in learning time, third in number of repetitions, and first in retention.

The oldest Subject, D, age 58, ranked fourth in learning time, fourth in number of repetitions, and eleventh in retention.

Subject A, the next to the youngest, age 21, ranked first in learning time, first in number of repetitions and second in retention. Really she ranked first if the average retention of the two methods of learning had been taken.

Subject F, age 42, ranked sixth in all three processes.

Certainly the results are encouraging to people of middle age.

In general, it appears that those who learn quickly and with few repetitions also retain well. This does not hold true with Subject D, probably because of the age factor; while Subject M seems to be an

example of the "slow but sure" type. Previous studies have shown that this result was to be expected.

As was expected also from the studies of Bennet[1] and others, those subjects who learned easily by one method, learned easily by the other method and vice versa. There is hardly an exception to this rule in the above results.

It seems, too, that meaningful matter, completely learned may, by an occasional repetition, be retained with a high degree of accuracy for a relatively long time. This is again shown by the large amount retained after seven days. This conclusion is very valuable for the student.

Definitely misleading results would have been obtained had group scores been used instead of individual scores. For example, the average number of repetitions for the whole group was less by the auditory method whereas, as a matter of fact, seven of the 12 subjects learned with fewer repetitions by the visual method. A glance at the wide variations in learning ability as shown by the tables, indicates the cause for this unreliability of group scores.

Conclusions

In answer to the original problem, it would appear from the above study that:

1. Neither method of presentation, auditory or visual, presents any marked degree of superiority in the rate of learning of meaningful, connected material.

2. Neither method has any distinct advantage over the other in regard to the number of repetitions required for learning.

3. Subsequent to the auditory method of presentation retention appears to be better than after the visual method of presentation.

It is further noted that, generally speaking:

4. The one who learns rapidly and with few repetitions also retains as well or better than the one who learns more slowly and with more repetitions.

5. The one who learns easily by one method also learns easily by the other method; and

6. The one who retains well through one method retains well through the other method.

7. Age does not greatly affect the ease of learning, but

[1] *Op. cit.*

8. Age does somewhat affect the ability to retain, though this effect is not as great as was expected.

9. A single repetition of the matter after one day, of well-learned material, so strengthens the associative processes that there is no marked decrease in the amount retained during the next day; and

10. A second repetition of the matter after two days assures a large retention for a relatively long time, in this case, one week.

11. Because of an insufficient number of subjects no conclusion could be drawn as regards sex differences.

12. Very wide differences between individuals are shown both in ease of learning, and in degree of retention.

13. Group scores in a study of this sort are unreliable.

It is realized that the main conclusion of this paper, viz.: that the auditory method of presentation is intrinsically superior for retention to the visual method of presentation, should be verified by further experimentation. It is particularly desirable that younger subjects should be employed. On the basis of these results alone, however, it would seem imperative that schools provide systematic training in auditory perception.

A STUDY IN THE PREDICTION OF HIGH SCHOOL SUCCESS

W. G. RECTOR

University High School, Oakland, California

There is evidence on every hand that efforts are being made to forecast the success of students in high school. Courses of study are being revised and adapted to the needs and capacities of students. The achievement of students is being checked against predicted success.

It is imperative, therefore, that we investigate very carefully the plans that are being used for prediction, for upon the validity of the criteria depends the practicability of any program for educational guidance.

Success in high school may depend upon a number of factors which may be determined before the students leave the elementary school. The difficult problem is to make a judicious selection of elements for prognosis. An easier problem, and one just as important as the selection of the elements, is the follow-up study to determine the validity of the criteria being used.

It is the purpose of this study to test the validity of a certain plan for the prediction of high school success. In the city in which this study was made the criteria used for prediction were:

1. Group intelligence quotients (GIQ) on the basis of scores on the Army Alpha Test.

2. Scholarship and application ratings on the basis of the average of three separate teachers' ratings. (These ratings were made on a seven-point scale while the students were in the high eighth grade.)

The above data were secured from the record cards of 431 tenth-grade students. The success of the 431 students was based on four final semester marks.

At the outset, the plan of attack was to show quantitatively the relationship between teacher-rated high school success and the three criteria for prediction. The method first employed was the percentage method. This method seemed to show that GIQ, scholarship, and application were valid means by which educational counselors could give guidance to freshman high school students.

As a test of the soundness of the results correlations between the variables were computed and found to be low. The percentage

method, so often employed to show causal relationship, had revealed the true situation with regard to mass data, but had utterly failed to show relationship for the individuals in the mass. The overlapping of grades made in high school by the different ability groups in GIQ, scholarship, and application was not shown by this method.

It became evident at once from the results secured by the percentage method that any claims made as to the validity of the factors used for prognosis have been based on follow-up studies using this same method. The point to be emphasized here is that the percentage method should *not* be used by research workers to determine the validity of criteria for *prediction*.

The validity, or better the lack of validity of GIQ, scholarship, and application as means for the prediction of high school success is shown by a study of Table I. The first column gives the subject; the second column the number of students taking the subject; the third, fifth and seventh columns show the correlation of high school success in the subject with GIQ, scholarship, and application respectively.

All of the correlations are low. The range is from $-.06$ between scholarship and shop success to $+.54$ between scholarship and mathematics. The correlation for all subjects combined and GIQ is .28; for scholarship .25; and for application .25.

These low correlations indicate that the criteria used for prediction are not valid. The correlation coefficients themselves constitute the major evidence for this conclusion; but in order to make clear their real significance, and, at the same time, to show the value they have for prediction purposes, 50 cases were selected from the total 431 students for special study.

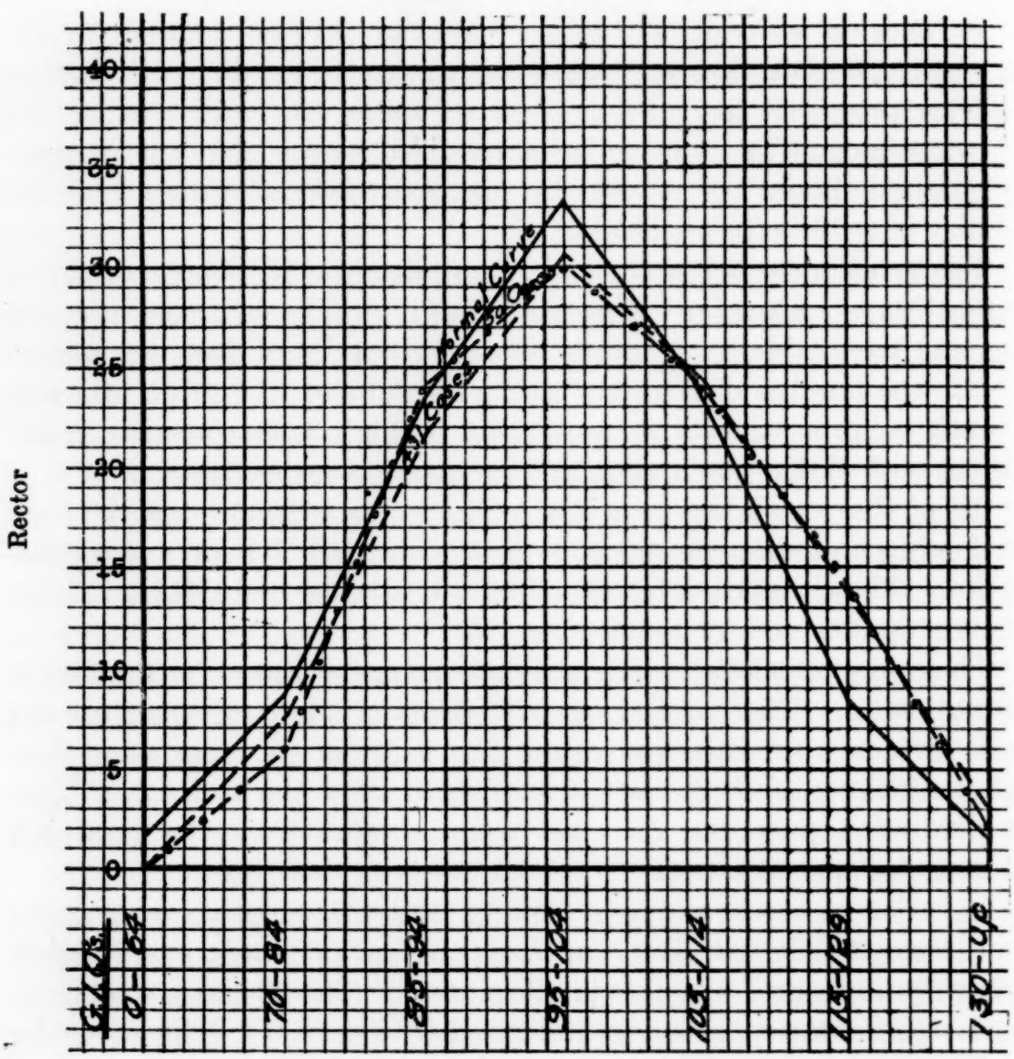
In order to secure a fair sampling, the 431 cases were arranged from lowest to highest GIQ, and from this arrangement every eighth and every ninth case was chosen alternately. The 50 cases, so chosen, are a representative group as indicated by Graph 1. The curve for the 50 cases is almost identical with that for the 431 cases. Furthermore, when tested for fairness of sampling by the formula

$$\sigma_s = \left[\frac{f\,(N - f)}{N}\right]^{\frac{1}{2}}$$

the greatest skewness from the normal curve, at 115–129 GIQ gives $\sigma_s = 1.96$. Whenever the actual difference is less than $3\sigma_s$ it is probable that this fluctuation is due to taking a small number of cases and

[1] See Rugg, H. O.: "Statistical Methods Applied to Education," p. 213.

GRAPH 1.—COMPARISON BETWEEN 50 GIQ's SELECTED FROM 431 GIQ's BY CHOOSING ALTERNATELY EVERY EIGHTH AND EVERY NINTH CASE WHEN ARRANGED IN ORDER FROM HIGHEST TO LOWEST CASE



	PER CENT	PER CENT	PER CENT	PER CENT	PER CENT	PER CENT	PER CENT
Normal	1.48	8.41	23.41	33.12	23.41	8.41	1.48
431 Cases	0	7.2	21.4	30.5	24.2	13.7	3
50 Cases	0	6	24	30	24	14	2

NOTE: Fairness of sampling is indicated by the close fit of the curves.

not due to constant causes which contribute to skewness.[1] The actual difference at 115–129 GIQ between the real number of cases and the theoretical number that should fall there is three, which is less than $3\sigma_s$. It is also evident that the greatest departure of the curve for 50 cases from the curve for 431 cases, at 85–94 GIQ, is less than the greatest departure from the normal curve at 115–129 GIQ; and hence, we are justified in assuming that the 50 cases shown are very closely representative of the 431 students.

TABLE I.—CORRELATION COEFFICIENTS BETWEEN SUCCESS IN HIGH SCHOOL AND GROUP INTELLIGENCE QUOTIENTS (GIQ), SCHOLARSHIP, AND APPLICATION AS CRITERIA FOR PROGNOSIS

Subject	Number of cases	GIQ		Scholarship		Application	
		r	PE	*r*	PE	*r*	PE
English	431	.25	.015	.31	.015	.27	.015
Mathematics	285	.28	.021	.54	.016	.25	.021
Science	237	.28	.027	.34	.026	.27	.028
Spanish	201	.15	.025	.24	.026	.21	.027
History	183	.33	.028	.40	.026	.34	.028
Home economics	153	.075	.037	.28	.034	.27	.034
Music	149	.29	.033	.24	.034	.25	.034
Drawing	147	.097	.034	.11	.034	.24	.032
Arithmetic	144	.33	.036	.35	.035	.40	.034
Typing	112	.21	.042	.13	.042	.26	.040
French	89	.092	.039	.19	.038	.12	.038
Bookkeeping	79	.22	.048	.46	.042	.34	.047
Latin	59	.38	.042	.54	.035	.49	.039
Shop	53	.13	.054	−.06	.055	.18	.053
Shorthand	49	.072	.068	.23	.068	.26	.065
High school success	431	.28	.0077	.25	.0077	.25	.0077
GIQ	431	...	.....	.44	.024	.26	.030
Scholarship	431	.44	.024	...	.....	.56	.022
Application	430	.26	.030	.56	.022		

NOTE.—High school success on the basis of four final semester marks for 431 students. GIQ on basis of scores on Army Alpha. Scholarship and application ratings based on the average of three separate teachers' ratings.

This table is read as follows: For English there are 431 cases. The correlation for GIQ is .25 with a probable error of .015; for scholarship *r* is .31, PE .015, etc.

[1] See Rugg, H. O.: "Statistical Method Applied to Education," p. 213.

The 50 cases chosen according to the above method are given in Table II.

For each case is given

Column 2, GIQ

Column 3, scholarship

Column 4, application

Column 5, average grade for all subjects combined

Column 6, sigma index[1] for GIQ

Column 7, sigma index for scholarship

Column 8, sigma index for application

Column 9, sigma index for high school success

Columns 10, 11, 12 respectively, prediction of high school success on the basis of GIQ, scholarship, and application[2]

---

[1] Sigma indices for any ability are obtained by dividing the deviation of the case from the mean of the group in that ability by the standard deviation of the whole group in that ability. By this method all abilities are reduced to a common denominator. Units which were directly incomparable are reduced to the same unit of measurement; and hence, relationship is clearly exhibited.

[2] The general formula for predicting one variable from another when the relationship (correlation) is known is given by

$$y = r\frac{\sigma_Y}{\sigma_X}x \pm \sigma_Y(1 - r^2_{XY})^{\frac{1}{2}}, \text{ where}$$

y = deviations of any score from the mean of the group Y
x = deviation of any score from the mean of the group X
r = the correlation between X and Y scores
σ_Y = standard deviation for the Y scores
σ_X = standard deviation for the X scores

The above formula may be written

$\frac{y}{\sigma_Y} = r\frac{x}{\sigma_X} \pm (1 - r^2)^{\frac{1}{2}}$, then we know that $r\frac{x}{\sigma_X}$ is the most likely $\frac{y}{\sigma_Y}$ of all individuals with the given $\frac{x}{\sigma_X}$ for this given correlation, r. $(1 - r^2)^{\frac{1}{2}}$ is the standard deviation (σ) of the column where $\frac{x}{\sigma_X}r$ is the mean of this column.

(The above formula, and the derivations are from Dr. R. H. Franzen.)

Adapting the above formula for the prediction of high school success, we have:

$g_i = r_{GI}\frac{\sigma_G}{\sigma_I}i \pm \sigma_G(1 - r^2_{GI})^{\frac{1}{2}}$, prediction of high school success on the basis of GIQ, where

g_i = prediction of the deviation from the mean of high school success on the basis of GIQ; and
i = deviation of GIQ from the mean of the group
r_{GI} = correlation between high school success and GIQ

Column 13, the most probable average grade that it is possible to predict for each student on the basis of GIQ, scholarship and application combined.[1]

Analysis of Table

Assuming that the reader has familiarized himself with the construction of Table I, the following interpretations will be easily understood:

Case 1.—A GIQ of 77 ranks this student the lowest of the group; and as interpreted in terms of an IQ would indicate feeble-mindedness. He is 1.9 sigma below the average of the group in GIQ (Column 6); .24 sigma below in scholarship (Column 7); and .56 sigma below in application (Column 8). On the basis of these data we are enabled to make a prediction of 2.83 ± .87[2] for success in high school on GIQ. Allowing for the standard error of estimate, ±.87, the range is from 1.96 to 3.70—from a college recommending grade to a grade indicating failure in high school (Column 10). For scholarship the prediction is 2.40 ± .88; or as above, 1.52 to 3.28 (Column 11); and for application 2.48 ± .88, or 1.60 to 3.36 (Column 12).

A combination of all three criteria for prediction gives a correlation with high school success of .34; and consequently slightly lowers the

---

σ_G = standard deviation of grades made in high school
σ_I = standard deviation of GIQ'S
$\sigma_G\ (1 - r^2_{GI})^{\frac{1}{2}}$ = the standard error of the prediction.

The most probable grade that it is possible to foretell, knowing the above relationships, is obtained by adding g_i to the average grade made in high school by the 431 students.

Formulas for predicting high school success on the basis of scholarship and application may be written similarly, as above.

[1] The regression equation by which the most accurate estimate of high school success can be made when influenced or partially dependent upon a set of three variables is as follows,

$$X_G = B_{GISA}\frac{\sigma_G}{\sigma_I}X_I + B_{GSIA}\frac{\sigma_G}{\sigma_S}X_S + B_{GAIS}\frac{\sigma_G}{\sigma_A}X_A + c,$$

in which the B's and the C are constants so derived that when employed as above a final grade (X_G) is obtained which is the most accurate estimate possible of attainment.

The above formula for a regression equation is adapted from Truman L.Kelley, "A Chart to Facilitate the Calculation of Partial Regression Coefficients of Correlation and Regression Equations," *Special Monograph, No.* 1 *of the Standard Educational Publications.*

[2] (+) deviations should be subtracted; (−) deviations added to obtain grade.

TABLE II.—TEACHER RATINGS OF HIGH SCHOOL SUCCESS IN COMPARISON WITH GIQ, SCHOLARSHIP, AND APPLICATION PREDICTIONS OF SUCCESS FOR 50 STUDENTS

1	2	3	4	5	6	7	8	9	10	11	12	13
Case	GIQ	Scholarship	Application	Average grade	$\frac{I}{\sigma^I}$	$\frac{S}{\sigma^S}$	$\frac{A}{\sigma^A}$	$\frac{G}{\sigma^G}$	Prediction of average grade on			Prediction for ISA combined
									GIQ (I)	Scholarship (S)	Application (A)	
1	77	C	C	2	−1.9	− .24	− .56	+ .38	1.96–3.70	1.52–3.28	1.60–3.36	2.22–3.38
2	80	C	C	1.87	−1.7	− .24	− .56	+ .53	1.90–3.64	1.52–3.23	1.60–3.36	2.18–3.34
3	82	C−	C	3.42	−1.5	−1.09	− .56	−1.18	1.86–3.60	1.71–3.41	1.60–3.36	2.20–3.36
4	85	C	C	1.83	−1.3	− .24	− .56	+ .57	1.80–3.54	1.52–3.28	1.60–3.36	2.11–3.27
5	86	C	C	2.54	−1.2	− .24	− .56	− .21	1.78–3.52	1.52–3.28	1.60–3.36	2.09–3.25
6	87	C	C−	3.86	−1.1	− .24	−1.41	−1.66	1.76–3.50	1.52–3.28	1.79–3.55	2.20–3.36
7	88	B	B	1.54	−1.1	+1.46	+1.14	+ .89	1.75–3.49	1.13–2.89	1.21–2.97	1.71–2.87
8	90	C	C	2.14	− .91	− .24	− .56	+ .23	1.71–3.45	1.52–3.28	1.60–3.36	2.03–3.19
9	91	C+	B	1.53	− .83	+ .62	+1.14	+ .90	1.69–3.43	1.33–3.09	1.21–2.97	1.72–2.88
10	92	C−	C−	3.00	− .76	−1.09	−1.41	− .71	1.67–3.41	1.71–3.41	1.79–3.55	2.18–3.34
11	93	C−	C	2.81	− .68	−1.09	− .56	− .51	1.65–3.39	1.71–3.41	1.60–3.36	2.04–3.20
12	93	B	B	1.36	− .68	+1.46	+1.14	+1.08	1.65–3.39	1.13–2.89	1.21–2.97	1.64–2.80
13	94	C	B	2.14	− .61	− .24	+1.14	+ .23	1.63–3.37	1.52–3.28	1.21–2.97	1.74–2.90
14	94	C+	C+	2.44	− .61	+ .62	+ .29	− .10	1.63–3.37	1.33–3.09	1.40–3.16	1.80–2.96
15	94	C+	C+	3.44	− .61	+ .62	+ .29	−1.20	1.63–3.37	1.33–3.09	1.40–3.16	1.80–2.96
16	96	C+	C+	1.80	− .45	+ .62	+ .29	+ .60	1.59–3.33	1.33–3.09	1.40–3.16	1.77–2.93
17	96	C+	B	2.44	− .45	+ .62	+1.14	− .10	1.59–3.33	1.33–3.09	1.21–2.97	1.68–2.84
18	97	C	B	3.00	− .38	− .24	+1.14	− .71	1.57–3.31	1.52–3.28	1.21–2.97	1.69–2.85
19	97	D	C+	3.29	− .38	−1.95	+ .29	− .70	1.57–3.31	1.92–3.68	1.40–3.16	1.92–3.08
20	98	D	B	1.94	− .30	−1.95	+1.14	+ .45	1.56–3.30	1.92–3.68	1.21–2.97	1.79–2.95
21	98	C+	B	2.00	− .30	+ .62	+1.14	+ .38	1.56–3.30	1.33–3.09	1.21–2.97	1.62–2.78
22	99	C+	C+	2.82	− .23	+ .62	+ .29	− .52	1.54–3.28	1.33–3.09	1.40–3.16	1.73–2.89
23	100	D	B	2.40	− .15	−1.95	+1.14	− .05	1.52–3.26	1.92–3.68	1.21–2.97	1.76–2.92
24	101	C	C	2.72	− .08	− .24	− .56	− .41	1.50–3.24	1.52–3.28	1.60–3.36	1.88–3.04

25	102	C	C	2.57	0	− .24	− .56	− .24	1.48–3.22	1.52–3.28	1.60–3.36	1.86–3.02
26	102	C	C−	2.49	0	− .24	−1.41	− .15	1.48–3.22	1.52–3.28	1.79–3.55	1.98–3.14
27	103	D	C−	3.00	+ .08	−1.95	−1.41	− .71	1.46–3.20	1.92–3.68	1.79–3.55	2.08–3.24
28	103	C	C	1.59	+ .08	− .24	− .56	+ .83	1.46–3.20	1.52–3.28	1.60–3.36	1.85–3.01
29	103	C+	C+	2.48	+ .08	+ .62	+ .29	− .14	1.46–3.20	1.33–3.09	1.40–3.16	1.67–2.83
30	104	C−	D	3.00	+ .15	−1.09	−2.30	−1.71	1.44–3.19	1.71–3.41	1.98–3.74	2.12–3.28
31	105	B	B	1.31	+ .23	+1.46	+1.14	+1.15	1.42–3.16	1.13–2.89	1.21–2.97	1.47–2.63
32	105	C+	B	1.94	+ .23	+ .62	+1.14	+ .45	1.42–3.16	1.33–3.09	1.21–2.97	1.52–2.68
33	106	C+	C+	2.78	+ .30	+ .62	+ .29	− .47	1.40–3.14	1.33–3.09	1.40–3.16	1.62–2.78
34	107	A	A	1.86	+ .38	+2.31	+2.00	+ .54	1.38–3.12	0.94–2.70	1.02–2.78	1.26–2.42
35	108	A	A	1.53	+ .45	+2.31	+2.00	+ .90	1.36–3.10	0.94–2.70	1.02–2.78	1.24–2.40
36	108	C+	B	2.14	+ .45	+ .62	+1.14	+ .23	1.36–3.10	1.33–3.09	1.21–2.97	1.47–2.63
37	110	C	C	1.67	+ .61	− .24	− .56	+ .75	1.33–3.09	1.52–3.28	1.60–3.36	1.75–2.91
38	111	C	C	1.81	+ .68	− .24	− .56	+ .59	1.31–3.05	1.52–3.28	1.60–3.36	1.73–2.89
39	112	C+	C+	2.18	+ .76	+ .62	+ .29	+ .19	1.29–3.03	1.33–3.09	1.40–3.16	1.54–2.70
40	112	B	C	2.76	+ .76	+1.46	− .56	− .45	1.29–3.03	1.13–2.89	1.60–3.36	1.61–2.77
41	113	C	B	2.60	+ .83	− .24	+1.14	− .27	1.27–3.01	1.52–3.28	1.21–2.97	1.46–2.62
42	114	C	C−	3.06	+ .91	− .24	−1.41	− .78	1.25–2.99	1.52–3.28	1.79–3.55	1.81–2.97
43	115	C	C	2.12	+ .99	− .24	− .56	+ .25	1.23–2.97	1.52–3.28	1.60–3.36	1.67–2.83
44	117	C+	B	1.90	+1.1	+ .62	+1.14	+ .50	1.20–2.94	1.33–3.09	1.21–2.97	1.34–2.50
45	118	C	C−	2.75	+1.2	− .24	−1.41	− .44	1.18–2.92	1.52–3.28	1.79–3.55	1.75–2.91
46	120	A	A	2.06	+1.4	+2.31	+2.00	+ .32	1.14–2.88	0.94–2.70	1.02–2.78	1.07–2.23
47	123	C	C	2.36	+1.6	− .24	− .56	+ .01	1.08–2.82	1.52–3.28	1.60–3.36	1.56–2.72
48	124	B	B	1.81	+1.7	+1.46	+1.14	+ .60	1.06–2.80	1.13–2.89	1.21–2.97	1.20–2.36
49	127	B	A	2.33	+1.9	+1.46	+2.00	− .02	1.00–2.74	1.13–2.89	1.02–2.78	1.03–2.19
50	132	C+	C+	1.5	+2.3	+ .62	+ .29	+ .93	0.91–2.65	1.33–3.09	1.40–3.16	1.25–2.41

range for prediction. The most probable prediction for all three criteria combined is 2.80 ±.58 (Column 13)—college recommending grade to a very mediocre grade in high school. His actual average grade for all subjects, however, is 2.00 (Column 5) .38 sigma (Column 9), above the mean of the group, contrary to what we would be led to expect from his standing in all elements for prediction. This grade is within the range for prediction on GIQ, scholarship, and application; but the wide range is little better than a mere guess.

Such an exposition as the above will, it is hoped, aid in the interpretation of low correlations such as those obtained in this study.

Case 2.—In a similar manner we may study Case 2. This student achieves a marked success in the academic subjects; but the elements for prediction indicate 2.76 with a standard error of estimate of ±.58, giving a range of 2.18 to 3.34.

Case 3.—The success that this student achieves shows a good positive correlation with the factors of prediction. If the sigma indices, in Columns 6, 7, 8, and 9 were exactly the same as regards sign and numerical value the relation would be the highest possible, and the correlation 1. Conversely, the greater the unlikeness the lower the correlation. The first two cases, it is evident, show a low negative correlation.

Case 4.—This case is very similar to Case 2, showing a negative correlation.

Case 5.—Again, we observe a low positive correlation with success —higher, however, than we would predict.

Cases 14 and 15.—The criteria for prediction for these two cases are exactly the same, and the prediction is 2.38 ± .58 in each case. One makes an average grade of 2.44, while the other makes an average grade of 3.44.

Turning to the latter part of the table:

Case 50 shows a positive correlation. This student makes an average grade comparable to the central tendency of the predictions.

Case 49 shows a decidedly negative correlation. The prediction is 1.61 ± .58; but the actual grade average is .33 point below the average for the very lowest GIQ (Case 1).

Case 48 again shows a positive correlation, while

Case 47 shows a decided negative correlation.

In a similar manner each case may be studied, and the significance of low correlations with consequent wide range for prediction exemplified.

Following is a brief summary of the findings:

Of the 18 students failing to make college recommending grades (grades below 2.49), only 9, or 50 per cent, could have been warned beforehand concerning their impending failure according to the criteria used for prediction (neglecting the PE in all cases).

The grades of 11 of the 18 students show $(+r)$ with GIQ
The grades of 7 of the 18 students show $(-r)$ with GIQ
The grades of 15 of the 18 students show $(+r)$ with scholarship
The grades of 3 of the 18 students show $(-r)$ with scholarship
The grades of 12 of the 18 students show $(+r)$ with application
The grades of 6 of the 18 students show $(-r)$ with application

No effort has been made to predict high school success for particular subjects, since all correlation coefficients are low, indicating the same general conclusions as given for all subjects combined; namely, that *GIQ*, *scholarship*, and *applicating* ratings are not valid for the prediction of high school success on the basis of teachers' marks.

AN EXPERIMENTAL INVESTIGATION OF THE CAUSES OF POOR SPELLING AMONG UNIVERSITY STUDENTS, WITH SUGGESTIONS FOR IMPROVEMENT[1]

FRIEDA A. KIEFER

Child Study Laboratory, Seattle, Washington
(Formerly Research Assistant, University of Michigan)
and

PAUL V. SANGREN
Western State Normal School, Kalamazoo, Michigan

University instructors are familiar enough with the seeming paradox that students of good to superior general ability not infrequently show very inferior ability in spelling. The explanation of this paradox is not as obvious as its existence.

Purpose

The purpose of this experiment was to determine the causes of poor spelling among university students of average to superior general ability and to test the efficacy of specific suggestions for the improvement of their spelling.

Our Experiments

Our investigation comprises three distinct experiments. These will be described in order.

Subjects

The first experiment was begun in the second semester of 1921–1922 at the suggestion of the Rhetoric Department of the University, which offered to send to us students who were found by the instructors in that Department to be conspicuously poor spellers. However, the students from the Rhetoric Department did not cooperate as fully as we expected, so a notice was published in the *Michigan Daily* urging

[1] This investigation was done under the general supervision of Dr. Guy M. Whipple, Professor of Experimental Education, University of Michigan. Dr. Clifford Woody, Professor of Education and Director of the Bureau of Educational Reference and Research, University of Michigan, gave many helpful suggestions. Professor Fred N. Scott and Mr. Lasher of the Rhetoric Department cooperated in helping us secure students for our work.

all students who experienced unusual difficulties in spelling to participate in our experiment. The students who responded to this invitation constituted our Experimental Group (hereafter referred to as Group E). The Control Group, good spellers (hereafter referred to as Group C), consisted of 13 members of a class in Mental Testing. The members of both groups varied in age, class, and university course. Both sexes were represented. The reader will gain a better idea of what we mean when we say that the members of Group E were "poor spellers," by glancing over the following list of misspellings of a few words, taken almost at random from a few similar ones:

ancient—anchent, anchant, anienct, acient
bouquet—becaucay, bouche, beauquet, beauque, boquet, beucay
calculation—calkilation, calation, cucalition
caliber—calibar, calabor, calabre, calibor, calaber, calabar, calabour
courtesy—courtize, curiscay, cortesy, curosy, curtesy
delicious—delious, deliscous, deliceous
fumigation—phermeagtion, fumagation, fumegation
ignorance—egarentc, egorance, ennorance, egorentc
nausea—nosocier, nosia, nausia, naucia, nauseau, naseau
quarantine—quarintine, quaranteen, quaranteed, quarentine

Experiment 1

The Tests.—The following list shows the experimental procedure followed for each student in Groups C and E:

Questionnaire on Personal and School History
Army Alpha Intelligence Test-Form 6
Questionnaire on Visual and Auditory Imagery (Betts)
Lowell Chart
Astigmatic Chart
Pilling-McCallie Audiometer
Dictaphone
Tachistoscope
Cancellation
Goddard Formboards
Ayres Spelling Lists "T" and "U"
Recognition Test
Recall Test
Reading Forward
Reading Backward
Writing
Foreign Words (Rote memory and rate of perception)
Marble Statue (Logical memory)

Space does not permit a complete description of each test, the apparatus used, the instructions given, the time occupied, and the method of recording.

Conclusions from Experiment 1

Our first experiment at Michigan yielded the following results:

1. Many of the poor spellers in our university group have defective vision, sometimes corrected, but often not.

2. None of them had defective hearing.

3. Careless or inaccurate observation of the written, printed, or heard word is characteristic of certain types of inefficient spellers.

4. Certain types of poor spellers have relatively short visual memory spans, especially for rote material.

5. Poor handwriting is noticeable only in a few cases. This is not of much importance, as many students use the typewriter.

6. Poor enunciation is typical of students having difficulty with spelling. This is quite evident with students whose misspellings are phonetic in character.

7. Temperamental traits are of primary importance in considering spelling ability. Most of the students were not careless or indifferent, but were not concerned enough about their poor spelling to make the effort necessary for improvement. They seemed to expect that we would furnish some magic cure for their defect.

8. Much poor spelling is due to bad spelling habits and to inadequate and poor training in the early school years.

9. Although there were no intellectually inferior students in either group, the poor spellers are relatively inferior when compared with the good spellers (r equals .54, PE equals .10). This is true also of their scholarship.

10. The temporary physical and mental conditions under which a student works, no doubt may account for many misspelled words.

11. In nine of the 14 cases studied, hereditary disability was reported, *i.e.*, similar disability appeared in parents or siblings.

12. The learning of shorthand has impaired the spelling efficiency of some of the poor spellers, according to their own claims. However, it is our opinion that the importance of this factor has been unduly emphasized.

13. There is practically no, or very little, relationship between types of imagery (visual, auditory, motor, etc.) and spelling ability.

Experiment 2[1]

The experiment begun in the second semester of 1921–1922, was continued throughout the first semester of the academic year (September, 1922 through January, 1923). While, in general, the

[1] Experiments 2 and 3 were done by Miss Kiefer assisted by Mr. Ethan Edloff, Instructor in Engineering English, who gave some of the tests and helped in the computations.

same plans were followed as in the previous investigation, a few changes and innovations were introduced. In addition to slight modifications in technique, the personnel of both groups was confined almost exclusively to freshmen in the College of Literature. A total of three or more hours of intensive training (not attempted previously) was given to practically all individuals in Group E. In some instances, where there was sufficient time and good cooperation, this drill was extended. To arouse enthusiasm and rivalry between the two groups, an oldfashioned "spelling match" concluded the experiment.

The Tests and Procedure

The questionnaire to determine types of Mental Imagery, the audiometer, the tachistoscope, List "T" of the Buckingham Revision of the Ayres Scale, the Recall Test, the Writing Test, and the Marble Statute Test were not used because they did not prove sufficiently diagnostic in Experiment 1.

The content of the Dictaphone Test was changed. A short paragraph containing words likely to be incorrectly pronounced, was substituted for the list of words used previously, because it was felt more natural effects would be obtained.

Another list of words which we called the University Test (because this was taken from words misspelled by university students) was substituted for the Recall Test in Experiment 1. The words were first presented in column form to Group E and Group C and the percentage of correct words calculated for each student. Group E was given extensive drill on these words. Three weeks later both groups were given the same words embodied in a completion test.

A Will-profile Test adapted to Personnel Work at the Carnegie Institute of Technology was given to all individuals in our second experiment.

Results and Their Interpretation

The results of our second investigation substantiate, in most instances, our conclusions reached in the first:

1. The most significant fact concerning our experiment is its practical value in demonstrating the possibility of improving poor spelling. As much as 36 per cent improvement was effected in one case. An inverse correlation (−.62, PE .09) obtained between original spelling

and amount of improvement made. That is, in general, the poorest spellers made the greatest gains in improvement.

2. Fewer individuals in both groups had defective vision. In the two cases where there was serious defect, the students had properly fitted glasses. One of these students, however, disliked to wear his glasses in public.

3. There was a closer relationship between spelling efficiency and efficiency in rote memory than in Experiment 1 (r equals .57, PE equals 10. This test was of value in cases E2, E3, E4, E6, E10, E11).

4. Poor enunciation was typical of most of the poor spellers.

5. The importance of temperamental traits was brought to our attention again. Of the 10 non-intellectual traits supposedly measured by the Will-profile Test, our results discriminate between the good and poor spellers in three, namely:

(*a*) "Perseveration" defined as continuous effort to attain an indefinitely defined end, shows quite a decided difference in favor of the good spellers (correlation between this test and spelling efficiency was .71, PE .068).

(*b*) "Speed of decision," *i.e.*, "quickness in reaching a decision or conclusion on objective matters," reveals a decided difference in favor of the good spellers (correlation between this test and spelling efficiency was .98, PE .025).

(*c*) "Freedom from inertia" or the tendency to work at one's highest speed without "external pressure" relatively was of more value in picking out the good spellers (r equals .44, PE 11).

Although tests designed to measure non-intellectual traits are regarded by some of our eminent psychologists as not yet perfected, the test mentioned has been found useful in personnel work and our results lead us to conclude that parts of it are valid in discriminating between good and poor spellers.

6. Good spellers under normal conditions are as a rule more rapid readers than poor spellers. They also surpass the poor spellers in making more rapid associations and groupings of letters in words. (Correlation between spelling efficiency and reading backward was .69, PE .067. Correlation between spelling efficiency and reading forward, first trial equals .54, PE .09; second trial .52, PE 10.)

7. Again the poor spellers are relatively inferior in intelligence when compared with the good spellers (r equals .46, PE 11). The correlation between spelling ability and scholarship was .64, PE .08. This again confirms our previous conclusion.

8. A high correlation obtained, naturally, between spelling efficiency and recognition of correct spelling (r equals .80, PE equals .05).

9. The individuals who rated high in spelling efficiency rated correspondingly high on the words given for recall (r equals .80, PE .05).

10. In our first experiment 9 of the 14 cases of poor spelling reported hereditary disability. We had no similar reports this year. One student ranking eighth, another ranking twenty-first, in spelling efficiency reported their mothers as having difficulty with spelling.

Experiment 3

All students in Group E of Experiment 2 were given three hours or more (some as many as 11 hours) of intensive drill in an effort to improve their spelling. The exact nature of the training varied with the individual needs. A brief sketch of some of the methods used in three cases ought to be suggestive.

Case E-6 (*Experiment* 2). *Time.*—One hour drill per week over a period of three weeks.

Outstanding Difficulty.—Pronunciation.

Methods.—(1) The "flash" method (as described later) was used.

2. The student profited considerably by "sounding out" the syllables.
3. Much *oral* spelling was given.
4. The following rules from Lester's Spelling Review were helpful:
   - Final e before a suffix.
   - Final consonant before a suffix.
   - Words ending in *ant, ent, ance, ence.*
   - Final y.
   - Rules on pronunciation.
5. Progress chart was kept.

Improvement 36.64 per cent.

Case E-10 (*Experiment* 2). *Time.*—One hour drill per week over a period of three weeks.

Outstanding Difficulty.—Pronunciation.

Methods.—(1) "Flash" method.

2. Difficult parts of words were capitalized instead of underscoring them.
3. A progress chart was kept.
4. "Ei" and "ie" rule (Alice) was learned.
5. Drill on words mispronounced (Lesson 6—Lester's Spelling Review).
6. Rules on pronunciation.
7. Rule on final consonant before suffix.

Improvement.—16.79 per cent.

Case E-14 (*Experiment* 1). *Time.*—One hour drill per week over a period of eleven weeks.

Outstanding Difficulties.—Poor visualization, difficulty in recognizing correct spelling and difficulty in pronunciation.

Methods.—(1) All words divided into syllables.

2. "Flash" method used.
3. Rules on:
   - Final consonant before suffix.
   - Final e before a suffix.
   - "Ei" and "ie" rule.
4. Distinguishing prefixes and suffixes.
5. Distinguishing per and pre.
6. Reviewed words in Experiment 1.
7. Reviewed 24 lessons from Lester's Spelling Review.

Improvement.—Considerable improvement was made but it was impossible to figure out the percentage as the words were not standardized according to their difficulty.

The "flash" method was used for all the students but was modified to individual needs.

In giving rules sometimes the investigator explained them step by step and gave examples, while often the initiative was put on the student. The student sometimes wrote on the blackboard, sometimes on paper and occasionally did both.

It was found helpful to use oral spelling in a few cases. In cases where there was extreme difficulty in pronunciation, the words were divided into syllables and "sounded out" syllable by syllable. In some cases exaggerated pronunciation was insisted upon at first until the student could pronounce the words with facility.

The students had many original ideas for mnemonic devices which proved quite useful.

Out of all of this intensive training and drill with these individuals, the following suggestions emerge as being the most useful we can present to those interested in training poor spellers among college students.

Suggestions for Improving Poor Spelling

1. *To increase the power of selective attention, use apperceptive methods.* Underline or write in colored letters the difficult portions of the word. If the student prefers, capitalize the difficult parts of the word instead of underscoring them.

2. *To develop greater carefulness and accuracy in observation use a variant of the "flash" method* (brief scrutiny of the word followed by an attempt to spell it and then verification). Write the misspelled words

on the blackboard or on paper. Underscore the portion of the word where the student's difficulty occurred. Tell the student to look at the word carefully, to visualize it, to pronounce it (emphasizing the points of difficulty), to spell it orally and then to write it without reference to the blackboard. Drill upon each word. Dictate the list of misspelled words until the student spells them perfectly. Give the words for recall after certain intervals.

3. *"Sound out" the syllables in using the "flash" method* for students whose misspellings are chiefly phonetic.

4. Make each student keep a progress chart on which he can figure out the percentage of errors made each week.

5. *Try oral spelling for students who have difficulty in pronunciation.*

6. For those students whose misspellings are chiefly phonetic, *use the dictaphone.*

At the University of Michigan the machines are available to those who desire to use them and are easy to manipulate. This is also an economical procedure. If students are properly directed, at least six or eight should be able to talk on one cylinder. Of course, the length of the material would obviously affect this. The instructor then must offer suggestions. Here the idea is, of course, for the student to hear a record of his own faulty enunciation and pronunciation. This method has proved very effective in the hands of instructors in certain departments of the University.

7. To aid pronunciation, *divide the word into syllables.*

8. *Do not hesitate to use mnemonic devices.* For instance, students who are poor in rote memory might associate the word "believe" with "Alice" remembering that the "e" follows the "i," etc. One student "piece" by "pie."

9. *Make every "effort to develop a spelling consciousness."* It is of vital importance. By this is meant that the student must become clearly conscious of his defects and definitely desirous of remedying them. Early spelling habits so long continued can only be broken through voluntary effort. To supply the student with an additional motive for improvement, give a few lectures on the cultural and economic value of spelling.

10. *Use any device that will stimulate the poor speller to concentrate his attention.* Helpful suggestions may be found in Professor Whipple's book on *How to Study Effectively.*

11. *Encourage the "dictionary habit."* Pocket editions are good investments.

12. Although the value of rules is questioned on account of the numerous exceptions, *note the likenesses of the different words.*

13. *Employ some spelling text adapted to help those in difficulty.* We found Lester's *Spelling Review* a most excellent text for our students. Explain and illustrate the rules used in this text, phrase by phrase. Ask the students to illlustrate them, also. Do not give more than one rule a week ordinarily. Do not insist upon the students committing the rules to memory.

14. Make a "profile" of each student on the various tests as in Case E-3.

Case E-3 (Experiment 1)

	Median Group C
1. Army Alpha Intelligence Examination	.....*
2. Range of Visual Attention	.....*
3. Cancellation Test (perceptual ability)	.....*
4. Reading Test (forward—first trial)	.....*
5. Reading Test (backward)	.....*
6. Logical memory-marble Statue	.....*
7. Rote Memory	.....*
8. Recognition Spelling Test (per cent correct)	.....*
9. Recall Spelling Test 1 (per cent correct)	.....*
10. Recall Spelling Test 2 (per cent correct)	.....*
11. Preliminary Spelling List (per cent correct)	.....*

Case E-3 (Experiment 1) shows graphically how a certain individual ranks in the various tests in comparison with the average performance of his own group and the control group. By a somewhat careful study of each case and a thorough analysis of the student's errors, together with temperament and attitude, it is possible to make a fair diagnosis and to suggest probable remedies.

While this experiment is in no sense exhaustive, it presents briefly a few suggestions and results obtained from one year of experimentation on 51 university students, two of whom remained in the work throughout the year.

It ought to be said that spelling deficiency arises from no single defect, so there is no standard diagnosis and no single remedy. In considering such a problem it is well to remember that there are all degrees of spelling efficiency and inefficiency, just as there are in other abilities. Since most causes of poor spelling can be eliminated, the most important thing is to find out the cause or causes and then to select the proper remedial measure. As pointed out above different

remedies apply to different types of defect. When possible, a detailed study of the students would yield the most satisfactory basis for diagnosis and prescription.

What we feel we have contributed is that on the one hand we find most atrocious spelling among students of superior ability and extended schooling, but on the other hand, it is often possible to locate the sources of the deficiency and by judicious and intensive training it is possible to effect a radical improvement.

NOTE ON PROFESSOR KELLEY'S FORMULA FOR DETERMINING THE SIGNIFICANCE OF DIFFERENCES

KARL J. HOLZINGER

University of Chicago

In the September 1923 issue of this journal Professor Kelley has set forth a formula for determining "the probable errors of our judgments of differences within the individual." The formula in his notation is written,

$$\text{PE (of individual } z_1 - z_2) = .6745\sqrt{2 - r_{1I} - r_{2II}} \qquad (1)$$

where z_1 and z_2 are "standard scores" on any two tests X_1 and X_2, and are defined by the equations,

$$z_1 = \frac{X_1 - M_1}{\sigma_1}, \text{ and } z_2 = \frac{X_2 - M_2}{\sigma_2}. \qquad (2)$$

The quantities r_{1I} and r_{2II} are reliability coefficients for tests X_1 and X_2 respectively.

Professor Kelley's proof of formula (1) by means of residuals is novel and interesting, but it is worthy of note that the same result may be obtained by formulas already known. The formula for the "probable error of response within an average individual"—as I prefer to call it—is familiar. For a given test X_1 this formula may be written,

$$\text{PE (of individual } X_1) = .6745\sigma \times \sqrt{1 - r_{1I}} \qquad (3)$$

A similar expression could, of course, be written for another test X_2. If the scores are given as "standard scores," as they must be in order to be comparable, formula (3) will be written,

$$\text{PE (of individual } z_1) = .6745\sqrt{1 - r_{1I}} \qquad (4)$$

and

$$\text{PE (of individual } z_2) = .6745\sqrt{1 - r_{2II}} \qquad (5)$$

Now if we assume response errors for z_1 and z_2 to be uncorrelated, we have by the usual formula $(\text{PE}_{a-b})^2 = (\text{PE}_a)^2 + (\text{PE}_b)^2$,

$$\text{PE (of individual } z_1 - z_2) = .6745\sqrt{2 - r_{1I} - r_{2II}} \qquad (6)$$

which is the same as (1). This formula is useful in determining whether or not a given difference $z_1 - z_2$ may be ascribed to response error of the individual.

We shall next derive a formula which is shorter than (1) and equivalent to it. Following Professor Kelley's notation the following list of symbols and their meaning is given;

z_1 and z_I = standard scores on first and second forms of X_1
z_2 and z_{II} = standard scores on first and second forms of X_2
r_{1I} and r_{2II} = reliability coefficients for X_1 and X_2 respectively
r_{12}, $r_{I\,II}$, r_{1II}, and r_{I2} = correlation coefficients between X_1 and X_2.
$d = z_1 - z_2$ = difference between scores on first forms.
$D = z_1 - z_{II}$ = difference between scores on second forms.
s and t = average or "true" scores of an individual on X_1 and X_2.
e_1 and e_I = individual response errors on X_1 by two forms.
e_2 and e_{II} = individual response errors on X_2 by two forms.

By use of the above definitions we may then write,

$$\left.\begin{aligned} z_1 &= s + e_1, \\ z_I &= s + e_I, \\ z_2 &= t + e_2, \\ z_{II} &= t + e_{II}. \end{aligned}\right\} \tag{7}$$

It is further evident that

$$d - D = Z_1 - Z_2 - Z_I + Z_{II} = (e_1 - e_2) - (e_I - e_{II}). \tag{8}$$

Now the result which we are seeking is the standard deviation of $(e_1 - e_2)$ or of $(e_I - e_{II})$. These last differences give the net response error of an individual score difference by first and second forms of tests X_1 and X_2. The standard deviation of these error differences for a group will give an approximation to the standard deviation of many similar differences for a single individual. This approximation is based on the assumption that the error differences for a group will distribute themselves in the same way as a set of similar differences for one individual. Inasmuch as this individual may be any one of the group, the above assumption does not seem unreasonable, and the approximation will be the best possible from two sets of measures for the group.

If we now assume that all of the errors, e, are uncorrelated, equation (8) gives the desired result at once. Remembering that the standard deviation of any z is equal to unity we may write,

$$\sigma^2_{d-D} = 4 - 2r_{12} - 2r_{1I} + 2r_{1II} + 2r_{I2} - 2r_{2II} - 2r_{I\,II} = \sigma^2_{e_1 - e_2} + \sigma^2_{e_I - e_{II}} \tag{9}$$

If the last two quantities be considered equal we have,

$$\sigma^2_{e_1-e_2} = \sigma^2_{d-D}/2 \tag{10}$$

or, $$.6745\ \sigma_{e1-e_2} = \text{PE (of individual } z_1 - z_2) = .477\ \sigma_{d-D} \tag{11}$$

Before commenting on this formula we shall next show that it is equivalent to (1). Returning to equation (9) and setting all four of the correlations between X_1 and X_2 equal, we have at once,

$$\sigma^2_{e_1-e_2} = 2 - r_{1I} - r_{2II} \tag{12}$$

The choice between formulas (10) and (12) will of course depend upon the constants required in a particular problem. If reliability coefficients are needed in addition to the above PE, then formula (12) will be preferable to (10) or (11). These reliability measures, however, are necessarily based on the particular group to which tests X_1 and X_2 are given and are therefore usually of little value in themselves, being affected by the age range and other peculiarities of the group. It would, of course, be a gross error to substitute in equation (12) reliability coefficients based on groups other than the one for which differences are to be tested. Thus since general reliability coefficients cannot legitimately be substituted in (12), and particular coefficients are often of little value, formula (11) will generally be the simpler and more direct equation to use.

To illustrate certain difficulties in the use of either formula, let us suppose that two forms of tests X_1 and X_2 have been given to the pupils in a single school grade with resulting reliability coefficients both equal to .5. Formula (1) then gives PE = .6745. If the difference $z_1 - z_2$ for pupil A is 1.2, we could write $z_1 - z_2 = 1.2 \pm .67$, and conclude that the observed difference might easily have arisen from response error. Now let us suppose that the same tests are given to the pupils in several grades (pupil A being included). The reliability coefficients this time might be both equal to .95. Formula (1) now gives PE = .21, so that for the same pupil A we could write $z_1 - z_2 = 1.2 \pm .21$ and conclude with almost certainty that the observed difference was not due to response error. It thus appears that the probable error of a given difference $z_1 - z_2$ may vary widely according to the number and "talent ranges" of pupils included with pupil A in its determination. In brief, formula (1) has the same defects as formula (4) which Professor Kelley himself criticizes elsewhere on the same grounds as those above.

In Professor Kelley's article cited at the beginning of this note he illustrates the use of formula (1) by pooling four Palo Alto eighth grade classes and working out reliability coefficients on the mixed

group. He then discusses the observed differences for four pupils. Thus for pupil D, "Difference: Science information—History and literature information = 1.7 standard deviations × PE of this difference = .4." From these and similar results he concludes that "very important differences may be taken as definitely established." Suppose, however, that Professor Kelley had calculated his reliability coefficients for pupil D and 95 of his fellow Palo Altos of the same age instead of grade. The resulting PE might then be twice as large, and the conclusion about differences for pupil D quite reversed.

Finally it may be noted that Professor Kelley has proposed a method for determining the "similarity" between two tests, or as we think of it, their effectiveness for differentiating pupils. The more "similar" the tests, the less effective for differentiating pupils. The ratio he employs may be written

$$\sqrt{\frac{2 - r_{1I} - r_{2II}}{2 - 2r_{12}}} = \sqrt{1 - r_{dD}} \qquad (13)$$

the right hand expression following at once from equations (7) and (10) if we set $\sigma_d = \sigma_D$.

In case such a measure of differentiation is a desirable one, the simpler formula would seem to be generally preferable. When $r_{dD} = 0$, we should have chance differences, the tests failing entirely to discriminate; when $r_{dD} = 1$, the discriminative capacity of the tests would be greatest, or according to Professor Kelley, they would be most "different."

By way of summary we may conclude that formula (1) is equivalent to applying formulas (4) and (5) to given differences, *i.e.*, it is the response error of an individual difference. The same results may also be obtained by the simpler formula (11). Finally, all of the above formulas have the defect of being functions of group reliability coefficients.

PROFESSOR KELLEY'S REPLY

Editor Journal of Educational Psychology

I wish to thank you for the opportunity to comment upon Dr. Holzinger's note in regard to the significance of differences within the individual. Dr. Holzinger concludes his note with the observation that "all of the above formulas have the defect of being functions of group reliability coefficients." This criticism is equivalent to the statement that "all of the above formulas yield different results dependent upon the range of talent examined." The methods of my article to be considered are:

(*a*) $d = x_1/\sigma_1 - x_2/\sigma_2$. The measure of difference within an individual.

(*b*) $\sigma_{d.\infty\omega} = \sqrt{2 - r_{1\text{I}} - r_{2\text{II}}}$. The standard error of the preceding measures.

(*c*) $d/\sigma_{d.\infty\omega}$. The measure of individual difference as a multiple of its standard error.

(*d*) $\sigma_d/\sigma_{d.\infty\omega}$. The standard deviation of the preceding, thus constituting a measure of group idiosyncrasy.

(*e*) $\sigma_d/\sigma_{d.\infty\omega}$. Preceding modified to represent the situation when reliabilities of the two tests are equal.

(*f*) The per cent of differences in individual test scores in excess of the chance percentage, derived from (*d*).

(*g*) The same measure as (*f*), but derived from (*e*).

A child's score as an abstract score is entirely meaningless. It is only when compared with some norm that it becomes significant. Different conclusions are drawn depending upon the norm; thus for example if compared with a fourth grade average we may conclude that it is superior to the average; if compared with a fifth grade norm that it is inferior, etc. The same observations apply with reference to differences. If a child receives a score of Y_1 in a reading test and Y_2 in an arithmetic test, superiority or inferiority of reading to arithmetic is not manifest—it only becomes so when a comparison with the records of some group is resorted to. This fact should not be looked upon as a short-coming, but rather as of the inevitable nature of things. Let us, in more detail, consider measures which are and should be interpreted in connection with a particular group, and those which have significance independent of a specific group.

Let lower case letters refer to the constants of a first group and to an individual's score when considered as belonging to this

group, and let capital letters have similar meanings with reference to a second group which, we will say, is composed of a wider range of talent. We have:

$\frac{x_1}{\sigma_1} - \frac{x_2}{\sigma_2}$ . . . . [The individual's superiority in Trait 1 to Trait 2 when measured in connection with the first group.]

$\frac{X_1}{\Sigma_1} - \frac{X_2}{\Sigma_2}$ . . . [The individual's superiority in Trait 1 to Trait 2 when measured in connection with the second group.]

These two values are clearly numerically unequal, but they have little significance as they stand. Thus, if we find that $(x_1/\sigma_1 - x_2/\sigma_2) = 1$, and accordingly say that the child is one standard deviation better in reading than arithmetic we have stated nothing of importance, for we do not know to what group is to be referred our measure of variability. If we do know the group, then the statement has meaning and becomes such a statement as, "the child is one fourth grade standard deviation better in reading than in arithmetic." It is of course true that the number of standard deviations that the child is better in reading than in arithmetic depends upon the spread of the group. It is obvious that such a statement with reference to quantities $(x_1/\sigma_1 - x_2/\sigma_2)$ is at all times necessary, but I find no more cause for criticising the measure on that account than I would for criticising a gross score because it must be referred to some norm before it becomes meaningful.

Two groups, for example, a fifth grade group and a 4-5-6-7 grade group will commonly have different standard deviations; but if a test is equally excellent in the narrow range as in the wide range, the standard error of estimate (given by $\sigma_{1.\infty} = \sigma_1\sqrt{1-r_{11}}$) will be the same in the two groups. We may utilize this common property of these two groups and obtain a method of interpretation which gives the same practical conclusion (identically the same if the reliabilities of the two tests are equal and substantially the same if unequal) no matter which group the pupil is placed in. This is the measure of idiosyncrasy lettered (*c*) above. The important property mentioned then is that $d/\sigma_{d.\infty\omega}$ for the narrow range is equal to $D/\sigma_{D.\infty\omega}$ for the wide range. The constancy of this ratio was first called to my attention by Dr. C. L. Huffaker in a letter from which he has kindly given me permission to quote as follows:

"1. If it can be assumed that as the true spread of a group increases in one trait it increases in another trait, at the same ratio, it can be shown that the per cent of measurable differences is constant regardless

of the true spread of the group and the necessary change in reliabilities of the measures used. In this case the per cent of measurable differences would be the correct method to be used in the interpretation of correlation coefficients."[1]

It thus happens that quite unlike the reliability coefficient, or Formula (4) of Dr. Holzinger's note which is a function of the reliability coefficient and of no other variable, the measure which I have used (*c*) is not affected by the range of talent involved, provided always that the test is equally excellent in the narrow and wide ranges considered.

In my article I derived a measure (*b*), from this ratio which I called "the percentage of differences in individual test scores in excess of the chance percentage." Obviously this also is unaffected by the range of talent of the group. There is, however, a very slight inexactness in this statement if the reliability coefficients of the two tests are unequal. In order to remedy this, one may use (*e*) which gives what the ratio $\sigma_{d.\infty\omega}/\sigma_d$ would equal provided the two tests were equally reliable, or (*g*), the percentage of differences in individual test scores in excess of the chance percentage in case the reliabilities of each test are equal, as was done on pages 332–333 of my article.

I would thus conclude that (*a*) and (*b*) are influenced by the range, of talent examined; that (*c*), (*d*) and (*f*) are not at all so influenced when the reliabilities of the tests for a given group are equal and but inappreciably so influenced when the reliabilities are unequal and such as are

---

[1] I have proved Dr. Huffaker's statement in case $r_{11} = r_{2\,11}$ but as I understand that he is to publish soon the recording of my proof here is not called for. I have also proved that if deviation values are divided not by the obtained standard deviations of the group but by the estimated true standard deviations and a measure of idiosyncrasy derived therefrom thus

$$\delta = \frac{x_1}{\sigma^{\infty}} - \frac{x_2}{\sigma^{\omega}}$$

leading to

$$\frac{\sigma^2 \delta}{\alpha^2 \delta.^{\infty\omega}} = \frac{\frac{1}{r_{1\,1}} + \frac{1}{r_{2\,11}} - 2r^{\infty\omega}}{\frac{1}{r_{1\,1}} + \frac{1}{r_{2\,11}} - 2}$$

we secure a ratio which is independent of the range of talent examined under certain important conditions, such as probably commonly prevail, even when r_{11} does not equal $r_{2\,11}$. This new measure (δ) though differing but slightly from (*d*) is probably a still more widely serviceable measure of idiosyncrasy.

usually found in the field of educational measurements; and finally that methods (*e*) and (*g*) are uninfluenced by the range of talent examined provided, as should be axiomatic, that the tests themselves are equally appropriate to the different ranges of talent examined. I consider these five measures: (*c*), (*d*), (*e*), (*f*), (*g*) to be applicable to a much wider field than Dr. Holzinger states and than I had surmised before receiving the observation of Dr. Huffaker.

TRUMAN L. KELLEY.

Stanford University

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS

To aid our readers who have depended upon this section for a bibliography of articles in current issues of other magazines, we here include those which have appeared since April 1, 1924.

Intelligence Testing

Facts, Inferences, and Assumptions in Intelligence Testing. George Frederick Miller. Educational Review, 1924, May, pp. 246–252. How the interpretation of these assumptions contributes to the present controversy about intelligence testing.

Psychological Examination of Pre-school Age Children. David Mitchell. The Pedagogical Seminary, 1924, June, pp. 108–146. A study of classification of over 2000 New York City children by use of the Stanford Revision of the Binet-Simon Tests.

On Family Resemblances beyond the First Degree of Relation. Emile S. Dexter. School and Society, 1924, April 26, pp. 501–502. A study of resemblance between second and third degree relations based on scholarship records and scores on intelligence tests.

The Nature and Validity of Subjective Estimates of Intelligence. David W. Oates. The Forum of Education, 1924, June, pp. 103–121. Estimates of intelligence by teachers compared with results on intelligence tests and school examinations.

Sorting the Students. Frank N. Freeman. Educational Review, 1924, November, pp. 169–174. Arguments for and against homogeneous grouping by ability and achievement tests.

The Intelligence of Chinese in Hawaii. Percival M. Symonds. School and Society, 1924, April 12, page 442. Results of intelligence, reading, vocabulary, completion and language tests on 513 Chinese children.

Some Results with the Stanford Revision of the Binet-Simon Tests. I. N. Madsen. School and Society, 1924, May 10, pp. 559–562. Results of the testing of 880 children in Idaho, including data on sex differences, grade placement, constancy of IQ, language difficulties and siblings.

Psychological Tests and Academic Salvage. Frank Herbert Palmer. Education, 1924, April, pp. 486–490. Emphasizes the need of encouraging all pupils impartially regardless of intellectual attainments.

The Measurement Movement—and the Man in the Street. Charles G. Reigner. Education, 1924, May, pp. 571–575. A criticism of tests of general intelligence as instruments of classification.

Correlation of the Grades in Practice-teaching Received by Seniors in a College for Teachers, with Their Scores in the Thurstone Group Intelligence Test. Hazel E.

Cooper. The Pedagogical Seminary, 1924, June, pp. 176–182. Shows that this test does not predict teaching ability.

The Primary Classification Test. L. W. Pressey. Journal of Educational Research, 1924, April, pp. 305–314. Discusses the revision of the Pressey Primer Scale.

Training Students in Group Intelligence Testing. Rudolph Pintner. Journal of Educational Research, 1924, April, pp. 271–280. The method of training by carrying out real testing in the schools.

Comparison of Certain Intelligence Scales. J. H. Wilson. The British Journal of Psychology, 1924, July, pp. 44–63. Comparisons of following group tests on 340 English children: Terman, Otis Advanced, National, Northumberland and Simplex.

The Measurement of Intelligence: The Present Status. E. L. Thorndike. Psychological Review, 1924, May, pp. 219–252. The first of a series of articles by Professor Thorndike on the measurement of intelligence.

An Evaluation of Some Information Questions. William R. Wilson, Gertrude Welsh and Harold Gulliksen. Journal of Applied Psychology, 1924, June, pp. 206–214. The selection of information questions for mental tests by the empirical method.

Psychological Tests and the Selection of Antioch Students. Iskander Hourwich. School and Society, 1924, August 2, pp. 156–160. Comparisons between students test scores and other measures of fitness for college admission used at Antioch.

Predicting Success or Failure in College at the Time of Entrance. J. B. Johnson. School and Society, 1924, June 28, pp. 772–776. A study of the value of psychological test scores and high school standing in predicting college success.

Predicting Success or Failure in College at the Time of Entrance II. J. B. Johnson. School and Society, 1924, July 5, pp. 27–32. Part II of a study of methods of prediction used in the University of Minnesota and comparisons with other schools.

The Slowing Up of Growth in Mental Test Ability. Hornell Hart. School and Society, 1924, November 1, pp. 573–574. A study of mental growth from the results of testing 579 children in Iowa with Army Alpha.

Correlations between Intelligence Quotients of Siblings. Hornell Hart. School and Society, 1924, September 20, page 382. A summary of three studies by the Iowa Child Welfare Commission on city children, rural children and university students.

Binet's Early Experiments on Intelligence Testing. Joseph Peterson. Peabody Journal of Education, 1924, September, pp. 87–92. The general content of an article in the *Revue Philosophique*, 1898, by Binet, on "Measurement in Individual Psychology."

A Psychological Study of Linguistic Abilities with Reference to the Result of Word Association Tests. Vivienne R. McClatchy and Mary Cooper. Journal of Experimental Psychology, 1924, October, pp. 371–381. A study of relations of the following tests: Free Association, Analogies, Proverbs and Sentence Completion.

The Constancy of the Intelligence Quotient. P. L. Gray and R. E. Marsden. The British Journal of Psychology, 1924, October, pp. 169–173. Results of retests after one, two and three years.

An Investigation of Group Intelligence Tests. R. R. Dobson. The British Journal of Psychology, 1924, October, pp. 162–168. Intelligence tests and selection pupils, limits of growth of intelligence and influence of environment.

On the Distribution of Intellect in the Ninth Grade. E. L. Thorndike and E. O. Bregman. Journal of Educational Research, 1924, November, pp. 271–278. Evidence that the distribution of ability in the ninth grade conforms very closely to the normal curve.

The Significance of Intelligence Tests in the University of Hawaii. Percival M. Symonds. School and Society, 1924, November 8, pp. 601–606. Summarized results of the use of intelligence tests in the University of Hawaii with recommendations for their proper use in college administration.

The Intelligence of Normal School Students and Its Relation to Learning and Teaching. I. N. Madsen. Educational Administration and Supervision, 1924, October, pp. 445–456. The relation of intelligence to success in normal school and to success in teaching.

An Evaluation of Some Group Intelligence Tests. John W. Rauth. The Catholic Educational Review, 1924, April, pp. 209–214. An evaluation of the Dearborn, Miller, Morgan, Otis and Terman Tests for high school use.

Educational Significance of Physical Status and of Physiological Mental, Emotional and Social Maturity. Arthur I. Gates. Teachers College Record, 1924, May, pp. 223–239. A discussion of current theories, summary of representative studies and results of an experiment in the Horace Mann School.

Scientific Method Applied to the Study of Individual Development. Edward A. Lincoln. School and Society, 1924, September 27, pp. 412–414. Some preliminary results of a comprehensive study at Harvard University of individual development by use of physical, mental and educational measurements.

Achievement Testing

Standard Tests and Their Use—a Symposium. E. L. Thorndike, Clifford Woody, M. R. Traube and William A. McCall. Teachers College Record, 1924, October, pp. 93–116. The evaluation of the use of tests, their present status and future expansion.

The Accuracy of Certain Standard Tests for School Classification. Percival M. Symonds. Journal of Educational Research, 1924, April, pp. 315–330. A study of the reliability, the difference in difficulty between two forms, and the gain due to practice effect of several widely used tests.

The Need of Instruments to Measure Kindergarten Education. Mary Maude Reed. Childhood Education, 1924, November, pp. 119–124. Discusses the need of achievement tests for measuring the results of kindergarten education with an evaluation of some present measures.

An Aspect of History Testing. Tyler Kepner. The Historical Outlook, 1924, December, pp. 414–416. The construction of a "Background Test in Social Sciences."

More Experiments with Scales. Carrie Barden. The English Journal, 1924, November, pp. 648–653. An experiment with the Hudleson Typical Composition scale in a Minnesota Teachers College.

A Method of Making Comparable Surveys of School Achievement. M. J. Van Wagenen. School and Society, 1922, June 14, pp. 707–712. A technique for comparing achievement scores in various subjects illustrated by a display of record cards.

Correlations between Intelligence and Scholarship. Gustave A. Feingold. The School Review, 1924, June, pp. 455–467. The relationship of intelligence scores, examination marks and recitation marks. A suggested method of grading achievement.

Mental Ability and the Examination. Robert T. Hance. School and Society, 1924, October 4, pp. 445–446. A comparison of examination grades with term marks for a large number of university students.

Relation between Rate and Quality. S. A. Courtis. Journal of Educational Research, 1924, September, pp. 110–131. A discussion of the problem of interpreting rate and quality scores.

The Measurement of Effort among High School Pupils. Gustave A. Feingold. Educational Administration and Supervision, 1924, September, pp. 385–395. Presents evidence that the EQ is a real measure of effort.

Educative Control by Means of a New Type of Measurement. John P. Herring. The Journal of Educational Methods, 1924, November, pp. 94–102. A report of a new type of measurement, the purpose of which is expressed as an "Instrument for the Educative Control of Behavior through Measurement and Methods."

The Use of Test Data as a Basis for Drill in Grammar. H. Ward McGraw. Journal of Educational Research, 1924, November, pp. 291–296. An experiment using the Kirby Grammar Test Scores to indicate needed drill.

A Latin Comprehension Test. B. L. Ullman and T. J. Kirby. Journal of Educational Research, 1924, November, pp. 308–317. Discusses the construction, reliability and validity of this test.

Method and Measurement in Spelling. Bertha M. Rogers. Journal of Educational Research, 1924, November, pp. 302–307. An experiment with a method designed to meet individual spelling needs.

A Test of Ability in the Pronunciation of Words. Arthur I. Gates. Teachers College Record, 1924, November, pp. 205–219. The construction, norms, reliability, validity and uses of the Gates Graded Word Pronounciation Test.

A Study of the Speed of Upper Grade Reading. C. E. Hulten. Journal of Educational Research, 1924, September, pp. 141–148. Indicates that speed may depend in part upon purpose and type of reading.

A Graphic Chart for Representing Educational Achievement Scores. M. R. Traube. Journal of Educational Research, 1924, May, pp. 411–414. Educational Achievement Chart, containing pictures of 14 thermometers each calibrated to represent the score from a different test.

The Reliability of Measurements in Spelling. R. L. Morton. The Journal of Educational Method, 1924, April, pp. 321–328. The report of a study to determine the accuracy with which the difficulty of a sampling of word from the Ayres Spelling Scale has been determined.

The Reliability of Teachers' Marks. Walter O. Shriner. Mathematics Teacher, 1924, November, pp. 426–443. A study based on marks in Algebra and English. Several past studies summarized.

The Educational and Accomplishment Quotients as an Aid in the Classification of Pupils. Marvin F. Beeson and Richard E. Tope. Journal of Educational Research, 1924, April, pp. 281–292. A study of classification in the schools of Grand Junction, Colorado.

The Reliability of Non-standardized Point Tests. L. O. McAfee. The Elementary School Journal, 1924, April, pp. 579–585. A comparison of the reliability of "simple associative reaction tests" yes-no tests and discussion tests.

A Variation in the "True and False" Achievement Test. George F. Miller. School and Society, 1924, August 23, pp. 250–251. An illustrative examination where student marks not only the true and false statements but also those which may be either true or false.

Written Examinations versus Standardized Tests. Walter S. Monroe. The School Review, 1924, April, pp. 253–265. A comparison of written examinations and standardized tests as to accuracy.

A First-grade Reading Test. Eliza F. Oglesby. Journal of Educational Research, 1924, June, pp. 29–41. Selection of materials, validity, and reliability of the Detroit Group Test in Word Recognition.

A Diagnostic Fractions Test. M. S. Robertson. Peabody Journal of Education, 1924, May, pp. 301–310. The construction and standardization of the Robertson Fractions Test.

The Status of Certain Basic Latin Skills. Leo J. Brueckner. Journal of Educational Research, 1924, May, pp. 390–402. Results of 26,297 Latin tests in the classical investigation—distributions and comparisons of four Latin tests.

The Effect of the Summer Vacation on the Reading Ability of First-grade Children. L. T. Brueckner and H. W. Distad. The Elementary School Journal, 1924, May, pp. 698–707. The differences in reading ability before and after the summer vacation as measured by standardized tests in Minneapolis.

The Reliability of Silent Reading Tests. Fowler D. Brooks. School and Society, 1924, May 31, pp. 652. Correlations between two trials of the Burgess, Courtis, Monroe and Haggerty Tests.

A New French Vocabulary Test. Alice M. Twigg. School and Society, 1924, December 6, pp. 734. A report of the compilation of a French vocabulary of 7261 words by word counts of French writings.

Learning

Types of Learning. Charles H. Judd. The Elementary School Journal, 1924, November, pp. 173–183. An analysis of the learning process in three fields, that of machine technique, the alphabet and the number system, emphasizing the need for a social background.

Persistence of Errors in Arithmetic. G. C. Myers. Journal of Educational Research, 1924. June, pp. 19–28. Points out that errors in number work are habits and once made, tend to be repeated.

Effect of Length of List upon Memory for Numbers. Edward S. Robinson and Chester W. Darrow. The American Journal of Psychology, 1924, April, 235–243. A study of length of list upon the curve of memorizing, difficulty of learning, amount retained and susceptibility to retroactive inhibition.

Should Poems Be Learned by School Children as "Wholes" or in "Parts." W. H. Winch. The British Journal of Psychology, 1924, July, pp. 64–79. Results of three experiments showing "part" learning to be decidedly superior.

The Warming-up Effect. Edward S. Robinson and William T. Heron. Journal of Experimental Psychology, 1924, April, pp. 81–97. An attempt to determine the conditions causing the warming up effect shown in some work curves.

A Theory of Learning. W. H. Pyle. Psychological Review, 1924, July, pp. 321–327. "A theory of brain action that will explain the various facts of the psychology of learning."

Repetition and Association in Learning. H. B. Reed. The Pedagogical Seminary, 1924, June, pp. 147–155. Experimental comparison of the value of learning materials made meaningful with those which are not.

An Experimental Isolation of Higher Level Habits. J. F. Dashiell. Journal of Experimental Psychology, 1924, October, pp. 391–397. A learning experiment in building up a hierarchy of habits on a higher level.

Criticisms of the Laws of Exercise and Effect. Hulsey Cason. Psychological Review, 1924, September, pp. 397–417. A criticism of these laws and an attempt to show that they do not fit the facts of learning.

The Memory Span of 600 Chinese School Children in San Francisco. Y. T. Hao, School and Society, 1924, October 18, pp. 507–510. A study of rate memory and its relation to race, sex and mental traits.

Memory. Edward S. Robinson. Psychological Bulletin, 1924, October, pp. 569–594. A review of memory studies since 1920 with bibliography.

The Value of Grouping Words According to Similar Difficulty in Spelling. W. F. Tidyman and Edith Johnson. Journal of Educational Research, 1924, November, pp. 297–301. A verification of C. A. Wagner's findings that grouping according to difficulty secures approximately 10 per cent better results.

Do Elementary School Pupils Know When They Make Mistakes in Spelling? W. F. Tidyman. School and Society, 1924, September 13, pp. 349–50. Study shows that pupils knew when they spelled correctly but were uncertain when they mispelled a word.

Mental Discipline and Fatigue

Transfer within a Narrow Mental Function. F. B. Knight and A. O. H. Selzafandt. The Elementary School Journal, 1924, June, pp. 780–787. A study of transfer in the addition of fractions.

The Present Status of the Theory of Transfer of Training. Brother Alphonsus, F. S. C. The Catholic Educational Review, 1924, June, pp. 338–345. Discussion of transfer on basis of past experimental studies.

Latin as a Preparation for French and Spanish. L. E. Cole. School and Society, 1924, May 24, pp. 618–622. Evidence of a small Latin-French and Latin-Spanish transfer.

Persistence, Success and Speed in a Mental Task. J. Crosby Chapman. The Pedagogical Seminary, 1924, September, pp. 276–284. The relationship of these factors in mental work and their corresponding effects on school tasks.

Some Effects of Prolonged, Unvaried Mental Work. F. M. Ritchie. The Forum of Education, 1924, June, pp. 83–98. A fatigue experiment with Kraepelin Multiplication Test by three individuals working in 10-hour periods.

Variations in the Mental Efficiency of Children during School Hours. Shepherd Dawson. The British Journal of Psychology, 1924, April, pp. 362–392. A study of the most effective placing of work periods.

Individual Differences

The Effect of Training on Individual Differences. H. B. Reed. Journal of Experimental Psychology, 1924, June, pp. 186–200. A study showing that equal

practice decreases individual differences and a criticism of the opposite conclusion reached by others.

Characteristic and Signification Differences between X and Z Pupils in the Detroit Pupil Schools. Anna M. Engel. The Elementary School Journal, 1924, June, pp. 747–754. Physical, mental, moral, pedagogical and social differences in the traits of children of superior and inferior intelligence.

Recording Individual Differences. Charles E. Neville. The Elementary School Journal, 1924, October, pp. 144–146. A suggested form for recording individual differences and its uses.

Sex Differences in Character as Indicated by Teachers' Ratings. Hornell Hart and Elmer Olander. School and Society, 1924, September 20, pp. 381–382. An Iowa Child Welfare Commission study. Girls rated more sincere, more earnest, more energetic and more kind.

An Experiment with A, B, and C Intelligence Groups. Irma Davis and Jenny Lind Green. The English Journal, 1924, September, pp. 457–462. A study of group differences in attacking an English project and in amounts of drill necessary to fix minimal facts in English.

The Ability of Men and Women to Judge Intelligence. L. W. Webb. School and Society, 1924, Aug. 23, pp. 251–254. The relative values of the judgments of men and women in rating the intelligence of the same people.

Comparison of White and Negro Children in Verbal and Non Verbal Tests. Dagne Summe. School and Society, 1924, April 19, pp. 469–472. A comparison of several thousand white and negro children on the National Intelligence Test and the Myers Mental Measure.

Exceptional Children

The Treatment of the Gifted Child in the Light of the Scientific Evidence. Frank N. Freeman. The Elementary School Journal, 1924, May, pp. 652–661. Points out that the real problem is not between acceleration and enrichment but between adjustment which aims to save time and that which aims to give opportunity to work at a higher intellectual level.

The Psychology of Superior Children. Albertine A. Richards-Nash. The Pedagogical Seminary, 1924, September, pp. 209–246. A study of superior children by means of achievement studies, case studies and biographical studies.

An Introduction to Biography for Young Children Who Test Above 150 *IQ.* Leta S. Hollingworth. Teachers College Record, 1924, December, pp. 277–287. An attempt to enrich the curriculum for intellectually gifted children through a study of biography.

Capitalizing the Extra-bright Child. Marguerite Aull. Peabody Journal of Education, 1924, November, pp. 141–146. Stresses the importance of developing the extra-bright child into a socially intelligent creature.

Classification for Instruction of Mentally Deficient and Retarded Children. J. E. Wallace Wallin. Mental Hygiene, 1924, July, pp. 753–768. Some considerations concerning the segregation of mentally handicapped children for instruction in special classes.

Educational Tests for Retarded School Children. A. Leila Martin and L. A. Pechstein. Journal of Educational Research, 1924, May, pp. 403–409. Evaluation of certain educational tests for use with retarded children.

My Experiment with Misfit High School Girls. Harriett Plimpton. Educational Review, 1924, November, pp. 178–180. The remedial teaching of a "slow progress" group in Philadelphia.

Case Studies of Reading Deficiencies in Junior High Schools. William S. Gray, Journal of Educational Research, 1924, September, pp. 132–140. Emphasizes the importance of detailed case studies, the conditions which justify such studies and steps in making them.

Diagnosing Individual Cases. J. Freeman Guy. Journal of Educational Research, 1924, September, pp. 100–109. The use of standardized tests and adequate records in diagnosis. Three illustrative cases cited.

Physical Defects—Their Relation to Achievement Scores and to Each Other. J. N. Mallory. Peabody Journal of Education, 1924, November, pp. 147–152. Data on the association of physical defects with mental and achievement test scores.

Educational and Vocational Guidance

An Individual Educational Guidance Card for College Students. Herbert A. Toops. School and Society, 1924, July 26, pp. 125–128. A form graphically showing a student's position on intelligence, achievement, age and health scales.

What Women in Colleges of Liberal Arts Specialize In. Agnes L. Rogers. School and Society, 1924, November 29, pp. 700–702. A report covering four years on the fields of specialization and their relation to intelligence scores.

Intelligence Tests as an Aid in the Diagnosis of Academic Maladies. C. E. Erffmeyer. School and Society, 1924, September 6, pp. 317–320. The use of intelligence tests in college work in predicting failures and in dealing with academic delinquency.

The Relation of IQ to the Prognosis of Special Class Pupils. August Dvorak. School and Society, 1924, May 21, pp. 736–744. A study of mental growth, showing that the normal child's IQ of 100 is constant but IQ's above 100 tend to increase and those below 100 tend to decrease.

Statistical Methods

The Influence of the Chance Imperfections of Measures upon the Relation of Initial Score to Gain or Loss. E. L. Thorndike. Journal of Experimental Psychology, 1924, June, pp. 225–232. Showing the need of a correction in correlating gain with initial ability when measures are unreliable.

A Formula to Correct for the Effect of Errors of Measurement on the Correlation of Initial Values with Gains. Godfrey H. Thompson. Journal of Experimental Psychology, 1924, August, pp. 321–324. A method of correcting correlations between initial value and gain when measures are unreliable.

A Note on the Correlation of Sums. El. Wynn Jones. The British Journal of Psychology, 1924, July, pp. 20–22. Formulas useful in correlating a single test with all tests pooled together.

The Sampling Error in the Theory of Two Factors. C. Spearman and K. Holzinger. The British Journal of Psychology, 1924, July, pp. 17–19. A correction for applying the correct criterion in the theory of two factors to abilities whose correlations are affected by errors of sampling.

Character and Personality

A New Method of Measuring Character Traits. Frank K. Shuttleworth. School and Society, 1924, June 7, pp. 679–682. A study of attitudes and interests toward wealth, demonstrating a method of measuring character traits.

A Report on Students' Attitudes toward Laboratory Work. Donald A. Laird. School and Society, 1924, May 17, pp. 589–592. A study of attitudes toward a laboratory course in psychology. Five out of every six students faked the experiments.

Moral and Social Development of the Six-year Old Child. Ellen A. Maher. The Pedagogical Seminary, 1924, September, pp. 268–275. The development of a tentative set of questions for measuring moral and social responses of six-year old children.

A Technique for Trait Analysis. W. W. Charters. Journal of Educational Research, 1924, September, pp. 95–100. A discussion of six problems in trait analysis namely: collection, definition, translation, condensation, evaluation and use.

The Downey Will Temperament Test in the Normal School. Arthur Kalstead. Journal of Educational Research, 1924, November, pp. 332–334. Correlation between the traits measured and grades in practice-teaching, education, art and penmanship.

Miscellaneous

General Review: Educational Psychology. Bird T. Baldwin. Psychological Bulletin, 1924, April, pp. 203–224. A review of the field since 1920, with extended bibliography.

The Determination of the Technical Vocabulary of the School Subjects. Luella Cole Pressey. School and Society, 1924, June 19, pp. 91–96. A detailed study of the technical vocabulary of the chief school subjects by word counts of over 100 different readers and 100 textbooks.

The Factor Concept in Education. S. A. Courtis. School and Society, 1924, April 12, pp. 413–423. Shows the importance of the law of the single variable in dealing with educational data.

Psychology as a Basis of Educational Methods. Charles H. Judd. The Elementary School Journal, 1924, October, pp. 102–112. A criticism of educational psychology which is based on individual instead of social psychology.

The Elimination of Children's Fears. Mary Cover Jones. Journal of Experimental Psychology, 1924, October pp. 382–390. A study of possible methods of eliminating children's fears. Two successful methods suggested.

Research Adventures in University Teaching. S. L. Pressey, L. W. Pressey and Helen Corbett. School and Society, 1924, November 15, pp. 635–638. An investigation of the amount of reading matter and methods of teaching for a course in educational psychology. Applicable also to other courses.

An Inquiry into the Teaching of Psychology in Training Colleges. Mrs. Susan Isaacs. The Forum of Education, 1924, June, pp. 133–140. Results of an inqiury from 63 Training Colleges and University Education Departments in Great Britain.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY JOHN HOCKETT

The Lincoln School of Teachers College

A Protest against the Current View of Instinct

Instinct; A Study in Social Psychology, by L. L. Bernard. New York: Henry Holt & Co., 1924. Pp. 550.

This volume is a milestone in the current movement to discredit the concept of instinct as it has been used in psychology for the past two decades. The author is professor of sociology in the University of Minnesota, and although he writes principally from the point of view of the sociologist, he shows a wide acquaintance with strictly psychological literature. The theory that much of human activity is instinctive, that "instinct dominates the formation of habits and determines the character of institutions," which was running at high tide when McDougall wrote his *Social Psychology*, is vigorously attacked in this book. The point of view is well summarized in the following statement: "The instincts are early overlaid by acquired habits in the process of adjusting the individual to his environment, and these habits are in turn overlaid by other tiers or stories of habit in which the native character of instinct ever constantly diminishes in proportion and intensity, until the child who has reached a rational age is reacting in nine-tenths or ninety-nine one hundredths of his character directly to environment, and only in the slight residual fraction of his nature directly to instinct." Professor Bernard points out that a solution of this question is important, as on the solution depends the direction of effort toward social amelioration. Using as a basis his thesis that much of human activity is not instinctively but environmentally determined he states as his belief that "an adequate control of social progress and of social and individual welfare lies mainly through the analysis and organization of the environment instead of through the control of heredity, although the latter is important."

The main part of the book is taken up with a determination of a definition of instinct which will be satisfactory to his opponents and

sanctioned by the most authoritative contributions to the subject. This is followed by an exhaustive survey of approximately 500 books and periodicals, representing slightly over 400 authors employing the concept of instinct, which disclosed no fewer than 849 separate types or variations of types of instincts and many classifications.

This book is a welcome resumé and summary of the present situation as regards instincts. It should be useful for collateral reading in general psychology classes and would serve as a text in courses on instincts and emotions.

The author, although attacking vigorously the position of the hereditarians, perhaps has failed to note that the hereditarians have shifted their emphasis from instincts to intelligence or general mental ability. Those who believe in the inheritance of differences in intelligence hold more strongly than ever their faith that through eugenics lies the key to control of ultimate social progress.

PERCIVAL M. SYMONDS.

Teachers College, Columbia University.

A SURVEY OF EXPERIENCE

The Field of Psychology: A Survey of Experience, Individual, Social and Genetic, by Madison Bentley. New York: Appleton and Co., 1924. Pp. 545.

This is not an attempt to give the reader an idea of the vast field over which modern psychology with all its various ramifications has spread, as one might at a first glance infer from the title of the book. On the contrary, the author keeps very rigidly to his own particular brand of psychology, which he tacitly assumes to be the only real psychology, and throws overboard with little or no argument a good many psychologies in more or less good standing at the present time. The author tells us that at the present time "Psychology is further from its moorings than it was at the end of the century," and evidently he regards it as his task to help to pull it back. But some of us may be glad that it is not now "moored" so closely to Wundt, Brentano, Stumpf, James and Ebbinghaus as it was, according to the author, at the end of the last century.

This dislocation, the author feels, has come about because of the employment of psychologists outside of their own domain, and the

increasing interest in the applications of psychology to education, medicine and commerce. All this has prevented the growth of a common point of view and has given rise to "partial" views, such as the clinical view of psychoanalysis, the behavioristic view of stimulus-and-response, or the efficiency view of intelligence. Part I of the book deals with "The Composition of Experience," an analysis of the constituents of experience, visual, auditory qualities, and the like. Part II is concerned with "The Organization of Experience," association, attention and habituation, which last is very briefly treated. Part III describes "The Psychosomatic Functions," perception, memory, emotion, and the like. Here under "the limits of psychosomatic function" a brief treatment of the measurement of intelligence is given, with much emphasis upon the crudeness of the methods and the lack of any consensus of opinion as to what intelligence is. Part IV is the last and shortest section of the book and deals with "The Socialization and Development of the Psychological Organism."

The book makes no particular appeal to students of educational psychology in general, and one has the feeling, whether rightly or wrongly, that the author has little interest in what he regards as a particular technological application of psychology. The book is well written and is altogether a scholarly piece of work.

R. PINTNER.

IMPLICATIONS OF ABNORMAL PSYCHOLOGY FOR EDUCATION

Abnormal Psychology and Education, by Frank Watts. New York: D. Appleton and Co., 1924. Third Edition. Pp. XIV + 220.

The author's intention in this volume has been to select facts and principles of abnormal psychology, which have implications for education, and to present them directly to teachers. The topics are organized under four general categories: (1) The Scope of Abnormal Psychology, (2) The Crowd at School: Its Control and Education, (3) Psychopathology and the Development of Personality, and (4) The Psychology of the Defective Mind: Its Influence upon Teaching Methods.

Abnormal psychology is defined as "the science which deals with those marked aberrations of the human mind, whether individual or social, and whether temporary or permanent, which do not make for

well-being." To try to cover a field so inclusive, discriminatingly, clearly and also very briefly, is to attempt the improbable. The table of contents promises a feast, but it is like the menu in a restaurant where too much is being undertaken. The fare is disappointing when actually set forth. Some of the necessary ingredients are lacking, parts seem half-baked, and worst of all, the flavor of certain courses is quite stale.

Of the four sections, that on crowd-management seems to the present reviewer to be the most helpful, suggesting what a teacher may expect from a class of children, as a collective unit. It is to be doubted, however, that the teacher will emerge from the brief sections on psychopathology and on intellectual deficiency, with ideas so firmly grasped as to be of service. There is too much plunging *in medias res*, with assumption of an assimilative background in the literature of psychology, which few teachers actually do possess.

The most serious adverse criticism is, however, that the volume has not been brought up to date. In the fields treated, a great number of early contributions have been superseded within the decade just passed. Most of the thought and evidence, on which the present edition is based, was current at least 10 years ago. This fault may be illustrated from the section which deals with intellectual deficiency.

As regards experimental education of the feebleminded, the author is back with Itard in 1802, and with Seguin, in 1866. As regards mental surveys and their results, he is still back with The Royal Commission of 1904. In mental measurement, despite citation of Terman, Yerkes, Pintner, Thomas, Burt and other recent investigators, he is really back with Binet and Simon, in 1908. One third of the section covering the topic of mental deficiency, is devoted to a detailed exposition of the educational theories and methods of Itard and Seguin, and of the application of their ideas by Montessori; while modern experimentation in special classes, and with industrial education and guidance, is not discussed.

The book will seem obsolete to those who profess educational psychology in the United States. They will wonder why a scholar apparently unfamiliar with so much of the research recently published, even with that in his own language, should think of revising his book.

LETA S. HOLLINGWORTH.

More Evidence on the Growth of Intelligence

A Study of the Upper Limits of the Development of Intelligence, by Florence M. Teagarden. New York: Teachers College, Columbia University, 1924. Pp. VI + 112.

Data concerning one of the prominent issues in present day psychology are contributed by this monograph. After a concise summary of the best contemporary thought upon the nature of intelligence, the shape of the curve of mental growth, the constancy of the IQ, the relation of age and variability of intelligence, and the age at which development ceases, this monograph reports a study of the latter problem. The subjects studied were some 400 persons ranging in age from 12½ to 20 years. Although the subjects were members of an institution, "Mooseheart," a home and school maintained by a fraternal order for children of deceased members, Dr. Teagarden insists that, due to the unusual nature of the institution, they are not atypical children and that their "home" life and training are not abnormal. These 400 young people were tested with one individual test, Stanford-Binet, and three group tests, Army Alpha, Pressey Senior Classification Test Form A, and Stenquist Mechanical Aptitude Tests, Test I.

The results of the 1600 test measures yield among others the following conclusions: that for the group studied (over 12½ years of age) there is a fairly even growth curve with negative acceleration; that within the group considered there is no cessation in growth of intelligence before 17½ or 18 years of age; and that variability as measured by individual tests increases with age.

A careful piece of research is set forth in this monograph. It should prove interesting to all students of the development of intelligence, and we hope may help to correct unjustified conclusions reached by some persons from the results of the army testing.

J. H.

The New Program in Mathematics

Teaching Junior High School Mathematics, by Henry C. Barber. Boston: Houghton Mifflin Co., 1924. Pp. 136.

The curriculum in mathematics in grades seven, eight and nine has, during the last 10 years, been extensively modified. The best evidence of this fact is the 8 or 10 series of textbooks in junior high school mathematics that have been published during this period. In these

texts there has been a revaluation of the aims of mathematical instruction. The first book on the teaching of junior high school mathematics discusses with unusual clarity and sanity the new attitude toward arithmetic, algebra and geometry in these grades. This little monograph, the first of a series of mathematical monographs being prepared under the editorial direction of Professor John W. Young, tells what the "new program" in mathematics is, why it should replace the "older program," and makes numerous suggestions on methods of teaching the newer materials. The discussion of the place of algebra and geometry in the seventh and eighth grades, the "new algebra" in the ninth grade, and approximate computation, interpret the spirit and purpose of junior high school mathematics most admirably. The author considers the purposes of mathematics in these grades to be the development of (1) a rational, problem-solving attitude, (2) skill in applying arithmetic to the every day problem of the home, store and farm, (3) the ability to compute with "approximate data," (4) an appreciation of, and some skill in using, the algebra formula and the equation, and (5) a knowledge of certain properties of space (intuitive geometry) that is to be acquired by the experimental laboratory method.

The section in which a hypothetical pupil summarizes his conception of the meaning and uses of seventh and eighth grade mathematics is an excellent concrete statement of the ideals of the new program in mathematics. If courses in mathematics are to be elective beyond the junior high school years, is it desirable to include in the ninth grade an introduction to demonstrative geometry? Mr. Barber is not certain, but he thinks *yes*. If time is not available for both numerical trigonometry and an introduction to demonstrative geometry, he prefers the trigonometry.

The monograph is particularly deserving of commendation for the success with which it interprets the meaning and purpose of the "new mathematics." The section on Computation will be read and re-read by teachers who desire to take a more common sense attitude toward computation in general, and who wish to familiarize themselves with the technique of Approximate Computation. It is in no sense a disparagement of the book to note that the author has given a little consideration to the methods of fixing skills or problems solving. The book will be very influential in helping both teachers and laymen understand and approve the type of mathematics which it describes.

JOHN R. CLARK.

Facts About the High School Principalship

The High School Principal—His Training, Experience, and Responsibilities, by Leonard V. Koos. Boston: Houghton Mifflin Co., 1924. Pp. XIII + 121.

This book reports the results of a questionnaire study of the high school principalship, contributed to by over 400 principals of schools well distributed as to location and size. It offers answers to such questions as these:

"What is the typical distribution by sex of principals in small and large high schools? What are typical salaries for the work, and is salary advancement characteristic of promotion from smaller to larger schools? What are the facts concerning the training of present incumbents—for example, as to its extent, degrees held, attendance in summer sessions, subjects emphasized, etc.? What aspects of their training are regarded as of greatest value by principals? What were their occupational plans as undergraduates? What have been their types of experience in educational work? Do they constitute a stable professional group, or do a majority plan to forsake the field at the first opportunity? What is the allowance of time for administrative and supervisory activities? Are principals generally given initiative in critical matters, or only in minor and clerical phases of the work?"

Besides specific data given in tables, graphs, and summaries, certain general conclusions are ventured. The high school principal is in the way of becoming respectably and significantly professional—is, in fact, already so in certain large schools of progressive communities. The things which he now chiefly lacks, and which he must increasingly seek, are better preparation, earlier vocational choice, the collection of more professional knowledge, and more responsibility in critical matters. In the meager literature now available concerning the high school principal, this little book most preeminently bears the impress of systematic, objective investigation. Its facts are important, and their presentation economical and effective. Most of it ought to be committed to memory by those now seeking promotion as principals, and by all who are interested in advancing the professional status of the occupation itself.

M. H. Willing.

Educational Problems with Answers

Principles of Education, by J. Crosby Chapman and George S. Counts. Houghton Mifflin Co., Boston, 1924. Pp. XVIII + 645.

This book is organized around question marks. Its four parts deal, respectively, with four major questions: What is the place of education in individual and social life? What are the psychological foundations of education? What are the sociological foundations of education? What principles govern the conduct of the school? Its 47 chapters deal with as many minor questions. Its two hundred or more sections within chapters deal with as many more questionlets. Finally, there are listed at the ends of chapters some four hundred inquiries of varying importance which the authors have not had time or space to answer. All this questioning and problem offering is doubtless good for the degree aspirant in education if his instructor sees to it that he feels himself interrogated; but it is somewhat irritating to the general reader who may want a less choppy, more obviously coherent discussion of such large and significant topics.

After all, though, the book does not merely raise questions. It answers them, and answers them with a clarity and an assurance that can come only from a fine mastery of the best speculative thought, empirical conclusion, and scientific research relating to the place, purpose and processes of education in the Great Society. It is not educational philosophy, but rather the interpretation and application of a philosophy to the problems of education. Assuming that education should be regarded as an instrument in human adjustment to a changing environment, the authors concern themselves with telling what that means in the light of social evolution, psychological fact, the six aspects of complete living—health, family life, economic life, civic life, recreational life, and religious life, and the existing institutional development of education.

The modernly trained school supervisor or administrator would find his task immeasurably lightened if teachers could be given the orientation provided by such a discussion as this. So much of educational philosophy at least seems badly needed as a requirement in the training of teachers. If, in somewhat different form, the same general point of view, the same blending of educational theories, the same deferential reference to thinkers, scientists, and keen observers of human society could be as effectively presented to the public at large, an even greater service might be rendered the cause of education.

M. H. Willing.